Some Imputation Methods in Double Sampling Scheme to Estimate the Population Mean

Narendra Singh Thakur¹, Kalpana Yadav² and Sharad Pathak³

¹Department of Mathematics and Statistics, Banasthali University, Banasthali, Rajasthan, INDIA
²Department of Statistics, Delhi University, Delhi, INDIA
³Department of Mathematics and Statistics, Dr. H.S. Gour Central University, Sagar, MP, INDIA

Available online at: www.isca.in, www.isca.me

Received 13th August 2013, revised 1st September 2013, accepted 10th October 2013

Abstract

In this paper, various imputation methods for missing values in double sampling scheme are suggested. Two different sampling designs in double sampling scheme are compared under imputed data. For different suggested estimators the bias and m.s.e up to the first order approximation are derived. Numerical study is performed over two populations using the expressions of bias and m.s.e and also efficiency compared with Ahmed estimators.

Keywords: Estimation, missing data, bias, mean squared error (m.s.e.), double sampling scheme, srswor, large sample approximation.

Introduction

Let us consider U = (1,2,3,...,N) be the finite population of size N and the character under study be denoted by y. Also, x be the ancillary variable which is highly correlated with study variable. If the population mean \( \bar{X} \) of the auxiliary variable x is unknown, then in such case the suggested estimator do not play satisfactory role in estimation. In such case the idea of two-phase sampling is helpful. A large preliminary simple random sample (without replacement) S of n units is drawn from the population on U and a second sample S of size n (n < N) is drawn in either following ways: i. the sample S is as a sub-sample from sample S (design I) as in figure 1, and ii. the sample S is independent to sample S without replacing S in the population (design II) as in figure 2.

Further, the sample S can be divided into two non-overlapping sub groups, i. the set of responding units, by R, and that of non-responding units by R. The number of responding units out of sampled n units be denoted by r (r < n).

For every unit \( i \in R \) \( y_i \) is observed, but for the units \( i \in R^c \), the \( y_i \) are missing and instead imputed values are derived. The ith value \( x_i \) of auxiliary variate is used as a source of imputation for missing data when \( i \in R^c \). Assume for S, the data \( x_i = \{ x_i : i \in S \} \) and for \( i \in S \), the data \( \{ x_i : i \in S \} \) are known with mean \( \bar{x} = \left( n \right) \sum_{i=1}^{n} x_i \) and \( \bar{y} = \left( n \right) \sum_{i=1}^{n} y_i \) respectively. The symbols that used are: \( \bar{x}, \bar{y} \) : the population mean of x and y respectively; \( \bar{y}, \bar{y} \) : the sample mean of x and y respectively; \( \bar{y}, \bar{y} \) : the sample mean squares of x and y respectively; \( C, C \) : the coefficient of variation of x and y respectively;

\[
\delta_1 = \left( 1 - \frac{1}{n} \right); \quad \delta_2 = \left( 1 - \frac{1}{n} \right); \quad \delta_3 = \left( 1 - \frac{1}{n} \right); \quad \delta_4 = \left( 1 - \frac{1}{n} \right); \quad \delta_5 = \left( 1 - \frac{1}{n} \right); \quad \delta_6 = \left( 1 - \frac{1}{n} \right); \quad f_i = \frac{r}{n},
\]

\[
E = \frac{(\delta_1 - \delta_2)(\delta_3 + \delta_4)}{\delta_2(\delta_3 + \delta_5 - \delta_3 + \delta_4)} \quad F = \frac{(\delta_1 - \delta_2)(\delta_3 + \delta_4)}{\delta_2(\delta_3 + \delta_5 - \delta_3 + \delta_4)} \quad G = \frac{(\delta_1 - \delta_2)(\delta_3 + \delta_4)}{\delta_2(\delta_3 + \delta_5 - \delta_3 + \delta_4)}.
\]
Large Sample Approximations

Let us consider \( \bar{y}_i = \bar{Y}(1 + e_i) \); \( \bar{x}_i = \bar{X}(1 + e_i) \); \( \bar{X} = \bar{X}(1 + e_i) \) and \( \bar{x} = \bar{X}(1 + e_i) \) . Now by using the concept of double sampling scheme and the mechanism of MCAR, for given \( r, n \) and \( n' \) we have:

<table>
<thead>
<tr>
<th>Designs</th>
<th>( E(e_i) )</th>
<th>( E(e_i^2) )</th>
<th>( E(e_i^3) )</th>
<th>( E(e_i^{*2}) )</th>
<th>( E(e_i^{*3}) )</th>
<th>( E(e_i^{*4}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \delta C_i )</td>
<td>( \delta C_i )</td>
<td>( \delta C_i )</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \delta C_i )</td>
<td>( \delta C_i )</td>
<td>( \delta C_i )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Designs</th>
<th>( E(e_i e_{i'}) )</th>
<th>( E(e_i e_{i'}) )</th>
<th>( E(e_i e_{i'}) )</th>
<th>( E(e_i e_{i'}) )</th>
<th>( E(e_i e_{i'}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>( \delta, \rho C_i C_x )</td>
<td>( \delta, \rho C_i C_x )</td>
<td>( \delta, C_i C_x )</td>
<td>( \delta, C_i C_x )</td>
<td>( \delta, C_i C_x )</td>
</tr>
<tr>
<td>II</td>
<td>( \delta, \rho C_i C_x )</td>
<td>( \delta, \rho C_i C_x )</td>
<td>0</td>
<td>( \delta, C_i C_x )</td>
<td>0</td>
</tr>
</tbody>
</table>

Proposed Strategies

Let \( y_j \) denotes the \( i^{th} \) observation of the \( j^{th} \) imputation strategy and \( b_1, b_2, b_3 \) are constants such that the variance of obtained estimators of \( \bar{Y} \) is minimum. We suggest the following tools of imputation:

\[
y_i' = \left\{ \begin{array}{ll}
\frac{y_i}{y_i} & \text{if } i \in R \\
\frac{1}{(1 - f_i)} \left[ k_i \left( \bar{x} - \bar{x}_i \right) + f_i \bar{x}_i \right] & \text{if } i \in R^c
\end{array} \right.
\]

under this strategy, the point estimator of \( \bar{Y} \) is

\[
t_i = \bar{y}_i + k_i \left( \bar{x} - \bar{x}_i \right) + f_i \bar{x}_i \quad ... (3.1)
\]

(3.1)

under this strategy, the estimator of \( \bar{Y} \) is

\[
ty_i = \left\{ \begin{array}{ll}
\frac{y_i}{y_i} & \text{if } i \in R \\
\frac{1}{(1 - f_i)} \left[ \theta_i \bar{x}_i + (1 - \theta_i) \bar{x} \right] - f_i & \text{if } i \in R^c
\end{array} \right.
\]

(3.2)

\[
ty_i = \frac{y_i}{\theta_i \bar{x}_i + (1 - \theta_i) \bar{x}} - f_i
\]

(3.3)

under this strategy, the estimator of \( \bar{Y} \) is

\[
ty_i = \left\{ \begin{array}{ll}
\frac{y_i}{y_i} & \text{if } i \in R \\
\frac{1}{(1 - f_i)} \left[ \theta_i \bar{x}_i + (1 - \theta_i) \bar{x} \right] - f_i & \text{if } i \in R^c
\end{array} \right.
\]

(3.4)
Hence the estimator of $\bar{Y}$ is

$$t_g = \frac{\bar{y}_r \bar{x}}{\theta_2 \bar{x} + (1-\theta_2) \bar{x}}$$  \hspace{1cm} (3.5)$$

$$y_{i_0} = \left\lfloor \frac{y_i - \bar{y}_r}{(1-f_i) \theta_1 \bar{x} + (1-\theta_1) \bar{x} - f_i} \right\rfloor$$ if \ \ \ \ \ \ \ \ \ i \in R$$

$$= \left\lfloor \frac{y_i - \bar{y}_r}{(1-f_i) \theta_1 \bar{x} + (1-\theta_1) \bar{x} - f_i} \right\rfloor$$ if \ \ \ \ \ \ \ \ \ i \in R^c$$  \hspace{1cm} (3.6)$$

Hence the estimator of $\bar{Y}$ is

$$t_{1_0} = \frac{\bar{y}_r \bar{x}}{\theta_3 \bar{x} + (1-\theta_3) \bar{x}}$$  \hspace{1cm} (3.7)$$

**Bias and M.S.E. of Proposed Methods**

Let $B(.)$ and $M(.)$, denote the bias and mean squared error (M.S.E.) of an estimator under a given sampling design $t = I, II$, then the bias and m.s.e of $t_7, t_8, t_9$ and $t_{1_0}$. The proofs of all these results are similar and therefore we will prove only one of them i.e. theorem 4.1.

**Theorem 4.1**: Estimator $t_7$ in terms of $e_i$; $i=1,2,3$ and $e_i$ could be expressed:

$$t_7 = \bar{Y}(1+e_1) + k_1 \bar{X}(e_3 - e_3) + k_2 \bar{X}(e_3 - e_2)$$  \hspace{1cm} (4.1)$$

by ignoring the terms $E[e_i e_i'], E[e_i e_i']$ for $r+s > 2$, where $r,s = 0,1,2,..$ and $i=1,2,3, j=2,3$ which is first order of approximation.

**Proof**: $t_7 = \bar{Y}(1+e_1) + k_1 \bar{X}(e_3 - e_3) + k_2 \bar{X}(e_3 - e_2)$

The estimator $t_7$ is an unbiased estimator under both the designs I and II i.e.

$$B[t_7]_I = 0$$  \hspace{1cm} (4.2)$$

$$B[t_7]_II = 0$$  \hspace{1cm} (4.3)$$

**Proof:**

$$B[t_7]_I = E[t_7 - \bar{Y}] = \bar{Y} - \bar{Y} = 0$$

$$B[t_7]_II = E[t_7 - \bar{Y}] = \bar{Y} - \bar{Y} = 0$$

The variance of $t_7$, under design I and II, up to first order of approximation could be written as:

$$V[t_7]_I = \delta_4 S_Y^2 + (\delta_2 - \delta_3)(k_1^3 S_X^2 - 2k_1 \rho S_Y S_X) + (\delta_1 - \delta_2)(k_2^3 S_X^2 - 2k_2 \rho S_Y S_X)$$  \hspace{1cm} (4.4)$$

$$V[t_7]_II = \delta_4 S_Y^2 + (\delta_3 + \delta_5)(k_1^3 S_X^2 - 2k_1 \rho S_Y S_X) + (\delta_4 - \delta_5)(k_2^3 S_X^2 - 2k_2 \rho S_Y S_X)$$  \hspace{1cm} (4.5)$$

**Proof:**

$$V[t_7] = E[ t_7 - \bar{Y}]^2 = E[ \bar{Y}^2 + k_1 \bar{X}(e_3 - e_3) + k_2 \bar{X}(e_3 - e_2)]$$

$$= E[ \bar{Y}^2 e_i^2 + k_1 \bar{X}^2(e_3 - e_3) + k_2 \bar{X}^2(e_3 - e_2)] + 2k_1 \bar{Y} \bar{X}(e_3 - e_3) e_i$$

$$+ 2k_2 \bar{Y} \bar{X}(e_3 - e_2) + 2k_2 \bar{Y} \bar{X}(e_3 - e_2) e_i$$

$$= E[ \bar{Y}^2 e_i^2 + k_1 \bar{X}^2(e_3^2 + e_3^2 - 2e_3 e_3) + k_2 \bar{X}^2(e_3^2 + e_3^2 - 2e_2 e_3)] + 2k_1 \bar{Y} \bar{X}(e_3 - e_3) e_i$$

$$+ 2k_2 \bar{Y} \bar{X}(e_3 - e_2) + 2k_2 \bar{Y} \bar{X}(e_3 - e_2) e_i$$

The International Science Congress Association

3
Under Design I (Using (4.6))
\[
V(\hat{\tau}_i) = \left[ \bar{Y}^2 \delta_1 C^2 + k_1 \bar{X}^2 \left( \delta_2 C^2 + \delta_3 C^2 - 2\delta_3 C^2 \right) + k_2 \bar{X}^2 \left( \delta_2 C^2 - \delta_3 C^2 - 2\delta_3 C^2 \right) \right] \\
+ 2k_1 \bar{Y} \bar{X} \left( \delta_3 \rho \delta_1 C_X - \delta_3 \rho \delta_1 C_X \right) + k_1 \bar{X} \left( \delta_3 \rho \delta_1 C_X - \delta_3 \rho \delta_1 C_X \right) \\
+ 2k_2 \bar{Y} \bar{X} \left( \delta_3 \rho \delta_1 C_X - \delta_3 \rho \delta_1 C_X \right)
\]
\[
= \left[ \bar{Y}^2 \delta_1 C^2 + k_1 \bar{X}^2 \left( \delta_2 - \delta_3 \right) + k_2 \bar{X}^2 \left( \delta_2 - \delta_3 \right) \right] \\
+ 2k_1 \bar{Y} \bar{X} \left( \delta_3 - \delta_2 \right) \rho \delta_1 C_X + 2k_2 \bar{Y} \bar{X} \left( \delta_3 - \delta_2 \right) \rho \delta_1 C_X
\]
\[
= \left[ \delta_1 S^2_Y + \left( \delta_2 - \delta_3 \right) k_1 S^2_X - 2k_1 \rho \delta_1 S_X \right] + \left( \delta_1 - \delta_2 \right) \left[ k_2 S^2_X - 2k_2 \rho \delta_1 S_X \right]
\]

Under Design II (Using (4.6))
\[
V(\hat{\tau}_i) = \left[ \bar{Y}^2 \delta_4 C^2 + k_1 \bar{X}^2 \left( \delta_2 C^2 + \delta_3 C^2 - 2\delta_3 C^2 \right) \right] \\
+ 2k_1 \bar{Y} \bar{X} \left( -\delta_3 \rho \delta_1 C_X \right) + 2k_1 \bar{X} \left( -\delta_3 \rho \delta_1 C_X \right) + 2k_2 \bar{Y} \bar{X} \left( -\delta_4 \rho \delta_3 C_X \right)
\]
\[
= \left[ \bar{Y}^2 \delta_4 C^2 + k_1 \bar{X}^2 \left( \delta_3 - \delta_4 \right) + k_2 \bar{X}^2 \left( \delta_3 - \delta_4 \right) \right] \\
+ 2k_1 \bar{Y} \bar{X} \left( -\delta_3 \rho \delta_1 \delta_4 S_X \right) + 2k_2 \bar{Y} \bar{X} \left( -\delta_4 \rho \delta_3 \delta_3 S_X \right)
\]
\[
= \delta_4 S^2_Y + \left( \delta_3 + \delta_4 \right) k_1 S^2_X - 2k_1 \rho \delta_3 \delta_4 S_X \right] + \left( \delta_4 - \delta_3 \right) \left[ k_2 S^2_X - 2k_2 \rho \delta_3 \delta_4 S_X \right]
\]

The minimum variance of the $\hat{\tau}_i$ is
\[
\left[ V(\hat{\tau}_i) \right]_{\text{Min}} = \left[ \delta_1 - \left( \delta_1 - \delta_3 \right) \rho^2 \right] S^2_Y
\]
\[
\left[ V(\hat{\tau}_i) \right]_{\text{II Min}} = \left[ \delta_4 - \left( \delta_3 + \delta_4 \delta_3 - \delta_3 \delta_3 \right) \delta_3 + \delta_3 \right] \rho^2 S^2_Y
\]

**Proof:**
First differentiate (4.4) with respect to $k_1$ and $k_2$, and then equate to zero, we get
\[
\frac{d}{dk_1} \left[ V(\hat{\tau}_i) \right] = 0 \Rightarrow k_1 = \frac{S_Y}{S_X} \quad \text{and} \quad \frac{d}{dk_2} \left[ V(\hat{\tau}_i) \right] = 0 \Rightarrow k_2 = \frac{S_Y}{S_X}
\]

After replacing value of $\beta_i$ in (4.4), we obtained
\[
\left[ V(\hat{\tau}_i) \right]_{\text{Min}} = \left[ \delta_1 - \left( \delta_1 - \delta_3 \right) \rho^2 \right] S^2_Y
\]

Similar to (i), we proceed for (4.5), we have
\[
\frac{d}{dk_1} \left[ V(\hat{\tau}_i) \right] = 0 \Rightarrow k_1 = \left( \frac{\delta_3}{\delta_3 + \delta_4} \right) \rho \frac{S_Y}{S_X} \quad \text{and} \quad \frac{d}{dk_2} \left[ V(\hat{\tau}_i) \right] = 0 \Rightarrow k_2 = \rho \frac{S_Y}{S_X}
\]
\[
\left[ V(\hat{\tau}_i) \right]_{\text{II Min}} = \left[ \delta_4 - \left( \delta_3 + \delta_4 \delta_3 - \delta_3 \delta_3 \right) \delta_3 + \delta_3 \right] \rho^2 S^2_Y
\]

**Theorem 4.2:** The estimator $\hat{\tau}_i$ in terms of $e_i, e_i, e_i, e_i$ and $e_i$ is
\[
\hat{\tau}_i = \bar{Y} \left[ 1 + e_i + \theta(e_i - e_i) + e_i e_i + (1 - 2\theta) e_i e_i + \theta e_i - (1 - \theta) e_i \right]
\]

The bias of the estimator $\hat{\tau}_i$ under design I and II respectively is
\[ B(\hat{t}_9) = \bar{Y}(\delta_1 - \delta_2 \left( \hat{\rho}_C \hat{C}_Y \right) ) \] (4.10)
\[ B(\hat{t}_8) = \bar{Y}(\delta_4 - \delta_5 \left( \hat{\rho}_C \hat{C}_Y \right) ) \] (4.11)

Mean squared error of \( \hat{t}_8 \) under design under I and II respectively is:
\[ M(\hat{t}_8)_I = \bar{Y}^2 \left[ \left( \delta_1 - \delta_2 \right) \left( \hat{\rho}_C \hat{C}_Y \right) \right] \] (4.12)
\[ M(\hat{t}_8)_I = \bar{Y}^2 \left[ \left( \delta_4 - \delta_5 \right) \left( \hat{\rho}_C \hat{C}_Y \right) \right] \] (4.13)

The minimum m.s.e. of \( \hat{t}_8 \) is
\[
\left[ M(\hat{t}_8) \right]_{min} = \left[ \delta_1 - \left( \delta_2 - \delta_5 \right) \right] \left( \hat{\rho}_C \hat{C}_Y \right) \] when \( \theta_1 = \frac{\hat{C}_Y}{\hat{C}_X} \) (4.14)
\[
\left[ M(\hat{t}_8) \right]_{min} = \left[ \delta_4 - \left( \delta_4 - \delta_5 \right) \right] \left( \hat{\rho}_C \hat{C}_Y \right) \] when \( \theta_1 = \frac{\hat{C}_Y}{\hat{C}_X} \) (4.15)

**Theorem 4.3:**

The estimator \( \hat{t}_9 \) in terms of \( e_1, e_2, e_3 \) and \( e'_3 \) is
\[ \hat{t}_9 = \bar{Y} \left[ 1 + e_1 + e_2 + e_3 + e'_3 - e_1e_3 - e_1e_3 + e_2e_3 + e_2e_3 + e_3e_3 \right] \] (4.16)

The bias of the estimator \( \hat{t}_9 \) under design I and II respectively is:
\[ B(\hat{t}_9)_I = \bar{Y}(\delta_2 - \delta_3 \left( \hat{\rho}_C \hat{C}_Y \right) ) \] (4.17)
\[ B(\hat{t}_9)_I = \bar{Y}(\delta_2 (\delta_3 + \delta_5) + \delta_3 \left( \hat{\rho}_C \hat{C}_Y \right) ) \] (4.18)

Mean squared error of \( \hat{t}_9 \) under design I and II respectively is:
\[ M(\hat{t}_9)_I = \bar{Y}^2 \left[ \left( \delta_1 - \delta_2 \right) \left( \hat{\rho}_C \hat{C}_Y \right) \right] \] (4.19)
\[ M(\hat{t}_9)_I = \bar{Y}^2 \left[ \left( \delta_4 - \delta_3 \right) \left( \hat{\rho}_C \hat{C}_Y \right) \right] \] (4.20)

The minimum m.s.e. of \( \hat{t}_9 \) is
\[
\left[ M(\hat{t}_9) \right]_{min} = \left[ \delta_1 - \left( \delta_2 - \delta_3 \right) \right] \left( \hat{\rho}_C \hat{C}_Y \right) \] when \( \theta_2 = \frac{\hat{C}_Y}{\hat{C}_X} \) (4.21)
\[
\left[ M(\hat{t}_9) \right]_{min} = \left[ \delta_4 - \delta_3 \left( \delta_1 + \delta_3 \right) \right] \left( \hat{\rho}_C \hat{C}_Y \right) \] when \( \theta_2 = \frac{\hat{C}_Y}{\hat{C}_X} \) (4.22)

**Theorem 4.4:** The estimator \( \hat{t}_{10} \) in terms of \( e_1, e_2, e_3 \) and \( e'_3 \) is
\[ \hat{t}_{10} = \bar{Y} \left[ 1 + e_1 + e_2 + e_3 + e'_3 - e_1e_3 - e_1e_3 + e_2e_3 + e_2e_3 + e_3e_3 \right] \] (4.23)

The bias of the estimator \( \hat{t}_{10} \) under design I and II respectively is:
\[ B(\hat{t}_{10})_I = \bar{Y}(\delta_2 + \delta_3 \left( \hat{\rho}_C \hat{C}_Y \right) ) \] (4.24)
\[ B(\hat{t}_{10})_I = \bar{Y}(\delta_2 (\delta_3 + \delta_5) + \delta_3 \left( \hat{\rho}_C \hat{C}_Y \right) ) \] (4.25)

Mean squared error of \( \hat{t}_{10} \) under design I and II respectively is:
\[ M_{(10)}^{(\cdot)} = \bar{Y}^2 \left[ \delta_1 C_1^2 + (\delta_1 - \delta_3) \left( \theta_2^2 C_2^2 - 2\theta_3 \rho C_3 C_X \right) \right] \]
\[ M_{(10)H}^{(\cdot)} = \bar{Y}^2 \left[ \delta_4 C_4^2 + (\delta_3 + \delta_4) \theta_2^2 C_2^2 - 2\theta_3 \delta_4 \rho C_3 C_X \right] \]  

(4.25)  
(4.26)  

The minimum m.s.e. of \( t_{10} \) is

\[ \begin{aligned} 
[M_{(10)}^{(\cdot)}]_{\text{min}} &= \left[ \delta_1 - (\delta_1 - \delta_3) \rho^2 \right] S_Y^2 \quad \text{when } \theta_3 = \frac{\rho C_Y}{C_X} \\
[M_{(10)H}^{(\cdot)}]_{\text{min}} &= \left[ \delta_4 - \delta_3^2 \left( \delta_3 + \delta_4 \right)^{-1} \rho^2 \right] S_Y^2 \quad \text{when } \theta_3 = \left( \frac{\delta_1}{\delta_3 + \delta_4} \right) \frac{\rho C_Y}{C_X} 
\end{aligned} \]

(4.27)  
(4.28)  

Comparisons

\[ \Delta_3 = \min [v(t_{10})] - \min [v(t_{11})] \]
\[ = \left[ \frac{1}{n} - \frac{1}{N} \right] S_Y^2 + \left[ \frac{2}{N} - \frac{2}{n} \right] \rho^2 S_Y^2 \\
\text{\( \left\{ t_{10} \right\} \) is better than } t_7, \quad \text{if } \Delta_3 > 0 \]
\[ \Rightarrow 2 \left[ \frac{1}{n} - \frac{1}{N} \right] \rho^2 < \left( \frac{1}{n} - \frac{1}{N} \right) \]
\[ \Rightarrow -2 < \rho < 2 \]

\[ \Delta_4 = \min [v(t_{10})] - \min [v(t_{11})] \]
\[ = \left[ \delta_1 - \delta_4 \right] S_Y^2 - \left[ \left( \delta_3 + \delta_5 \right)^3 \delta_5^2 + (\delta_4 - \delta_3) \right] \rho^2 S_Y^2 \\
\text{\( \left\{ t_{11} \right\} \) is better than } t_7, \quad \text{if } \Delta_4 > 0 \]
\[ \Rightarrow \rho^2 \left( \frac{\delta_1 - \delta_4}{\delta_3 + \delta_5} \right) < \left\{ \delta_3^2 + (\delta_4 - \delta_5) \left( \delta_1 + \delta_5 \right) \right\} \]
\[ \Rightarrow -E < \rho < E \]

\[ \Delta_5 = \min [v(t_{11})] - \min [v(t_{11})] \]
\[ = \left[ \frac{1}{n} - \frac{1}{N} \right] S_Y^2 \]
\text{\( \left\{ t_{11} \right\} \) is better than } t_8, \quad \text{if } \Delta_5 > 0 \]
\[ \Rightarrow \left[ \frac{N - n}{nN} \right] > 0 \Rightarrow N - n > 0 \Rightarrow n' < N \]

which is always true.

\[ \Delta_6 = \min [v(t_{11})] - \min [v(t_{11})] \]
\[ = \left[ \frac{1}{n} - \frac{1}{N} \right] S_Y^2 + \left[ \frac{2}{N} - \frac{2}{n} \right] \rho^2 S_Y^2 \]

\[ \text{\( \left\{ t_{11} \right\} \) is better than } t_8, \quad \text{if } \Delta_6 > 0 \]
\[ \Rightarrow \left( \frac{N - n + n'}{N(N - n)} \right) > 0 \Rightarrow n' > 0 \]

\[ \Delta_7 = \min [v(t_{11})] - \min [v(t_{11})] \]

International Science Congress Association
\( \left( t'_{17} \right) \) is better than \( t_9 \), if \( \Delta_{17} > 0 \)  
\[ \Rightarrow \rho^2 < \frac{1}{2} \quad \Rightarrow -\frac{1}{2} < \rho < \frac{1}{2} \]

\[ \Delta_{18} = \min \{ \hat{V}(t_9) \} - \min \{ \hat{V}(t'_{18}) \} = \left[ \delta_{14} - \delta_1 \right] S_y' - \left[ \delta_{15} - \left( \delta_3 + \delta_4 \right) \right] \rho^2 S_y \] 
\( \left( t'_{18} \right) \) is better than \( t_9 \), if \( \Delta_{18} > 0 \)

\[ \Rightarrow \rho^2 < \left( \frac{\delta_{14} - \delta_4}{\delta_3 + \delta_4} \right) \quad \Rightarrow -F < \rho < F \]

\[ \Delta_{19} = \min \{ \hat{V}(t_{10}) \} - \min \{ \hat{V}(t'_{19}) \} = \left[ \frac{1}{n} - \frac{1}{N} \right] S_y' + \left[ \frac{2}{n} - \frac{2}{N} \right] \rho^2 S_y' \] 
\( \left( t'_{19} \right) \) is better than \( t_{10} \), if \( \Delta_{19} > 0 \)

\[ \Rightarrow \rho^2 \cdot \left( \frac{1}{n} - \frac{1}{N} \right) < \left( \frac{1}{n} - \frac{1}{N} \right) \]  
\[ \Rightarrow -\frac{1}{2} < \rho < \frac{1}{2} \]

which is always true.

\[ \Delta_{20} = \min \{ \hat{V}(t_{16}) \} - \min \{ \hat{V}(t'_{20}) \} = \left[ \delta_{16} - \delta_2 \right] S_y' - \left[ \delta_{17} - \left( \delta_3 + \delta_4 \right) \right] \rho^2 S_y' \] 
\( \left( t'_{20} \right) \) is better than \( t_{16} \), if \( \Delta_{20} > 0 \)

\[ \Rightarrow \rho^2 < \left( \frac{\delta_{16} - \delta_4}{\delta_3 + \delta_4} \right) \quad \Rightarrow -G < \rho < G \]

**Numerical Illustrations**

We consider two populations A and B, first one is the artificial population of size \( N = 200 \) [source Shukla and Thakur (2008)]\(^5\) and another one is from Ahmed et al. (2006)\(^6\) with the following parameters:

<table>
<thead>
<tr>
<th>Table-1</th>
<th>Population Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N )</td>
</tr>
<tr>
<td>A</td>
<td>200</td>
</tr>
<tr>
<td>B</td>
<td>8306</td>
</tr>
</tbody>
</table>

Let \( n = 60, n = 40, r = 5 \) for population A and \( n = 2000, n = 500, r = 15 \) for population B respectively. Then the bias and M.S.E of suggested estimators under design I and II (using the expressions of bias and m.s.e. of Section 4) and Ahmed et al. (2006) methods (see Remark-1) are given in table 2, 3 and 4 for population A and B respectively.

<table>
<thead>
<tr>
<th>Table-2</th>
<th>Bias and MSE for Population – A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimators</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
</tr>
<tr>
<td>( t_7 )</td>
<td>0</td>
</tr>
<tr>
<td>( t_8 )</td>
<td>-1.40126E-06</td>
</tr>
<tr>
<td>( t_9 )</td>
<td>2.66906E-08</td>
</tr>
<tr>
<td>( t_{10} )</td>
<td>-.025405</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table-3</th>
<th>Bias and MSE for Population – B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimators</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
</tr>
<tr>
<td>( t_7 )</td>
<td>0</td>
</tr>
<tr>
<td>( t_8 )</td>
<td>0.00000381</td>
</tr>
<tr>
<td>( t_9 )</td>
<td>0.00000006</td>
</tr>
<tr>
<td>( t_{10} )</td>
<td>-.34747</td>
</tr>
</tbody>
</table>
Table 4

| Estimators | Population A | | | Population B | | |
|------------|-------------|---|---|-------------|---|
|            | Bias        | MSE | Bias | MSE         | |
| $t_7$      | 0           | 9.759633 |  | 0           | 16358.62 |
| $t_8$      | -0.000595   | 12.73984 | -0.09258 | 16531.89 |
| $t_9$      | -0.000068   | 35.83645 | -0.09527097 | 22319.77 |
| $t_{10}$   | -0.000663   | 9.759633 | 0.095271 | 16358.62 |

The sampling efficiency of suggested estimators under design I and II over Ahmed et al. is defined as:

$$E_i = \frac{Opt[M(i)]}{Opt[M(i)^t]}; \quad i = 7, 8, 9, 10; \quad j = I, II$$  \hspace{1cm} \text{…(*)}$$

The efficiency for population A and B respectively given in table-5.

Table 5

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Design I</th>
<th>Design II</th>
<th>Design I</th>
<th>Design II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population A</td>
<td></td>
<td>Population B</td>
<td></td>
</tr>
<tr>
<td>$E_1$</td>
<td>1.118298</td>
<td>3.967027</td>
<td>0.996435</td>
<td>1.374513</td>
</tr>
<tr>
<td>$E_4$</td>
<td>0.817709</td>
<td>0.966518</td>
<td>0.992239</td>
<td>0.999219</td>
</tr>
<tr>
<td>$E_9$</td>
<td>0.985928</td>
<td>1.026349</td>
<td>0.974671</td>
<td>1.000879</td>
</tr>
<tr>
<td>$E_{10}$</td>
<td>0.948329</td>
<td>1.156977</td>
<td>0.965441</td>
<td>1.001553</td>
</tr>
</tbody>
</table>

Remark-1: Under the setup when the population mean is known of auxiliary variable is known Ahmed et al. (2006) proposed some imputation methods and derived their properties. From which authors are discussing with four methods of them for comparison purpose. Let $y_i$ denotes the $i^{th}$ available observation for the $j^{th}$ imputation and $k_i, \ i = 1, 2$ and $\theta_i, \ i = 1, 2, 3$ is a suitably chosen constant, such that the variance the resultant estimator is minimum. Imputation methods are:

$$y_{ni} = \begin{cases} y_i & \text{if } i \in R \\ \bar{y}_r + \frac{nk_i}{(n-r)}(\bar{x} - \bar{y}) + k_i(x_i - \bar{x}) & \text{if } i \in R^C \end{cases} \quad \text{(6.1)}$$

Under this method, the point estimator of $\bar{Y}$ is

$$t_7 = \bar{y}_r + k_1(\bar{x} - \bar{y}) + k_2(x - \bar{x}) \quad \text{(6.2)}$$

Lemma 1: The bias, variance and minimum variance at $k_1 = k_2 = \frac{S_{xy}}{S_x^2}$ of $t_7$ is given by

$$B[t_7] = 0 \quad \text{(6.3)}$$

$$V(t_7) = \left(1 - \frac{1}{N}\right)S_Y^2 - 2S_{xy}\left[k_1\left(1 - \frac{1}{N}\right) + k_2\left(\frac{1}{r} - \frac{1}{n}\right)\right] + S_x^2\left[k_1\left(1 - \frac{1}{N}\right) + k_2\left(\frac{1}{r} - \frac{1}{n}\right)\right] \quad \text{(6.4)}$$

$$V(t_7)_{\min} = \left(1 - \frac{1}{N}\right)S_Y^2(1 - \rho^2) \quad \text{(6.5)}$$

$$y_{ni} = \begin{cases} y_i & \text{if } i \in R \\ \bar{y}_r \left(\frac{x_i + \frac{r}{n-r} \bar{x}}{\theta_1 x + (1 - \theta_1) \bar{x} - \frac{r}{n-r} y_i} \right) & \text{if } i \in R^C \end{cases} \quad \text{(6.6)}$$
Under this method, the point estimator of $\bar{Y}$ is

$$t_8 = \frac{\bar{y}_r \bar{x} - \bar{y}_r C^2 Y}{\theta_1 \bar{x} + (1 - \theta_1) \bar{x}}$$

(6.7)

Lemma 2: The bias, mean squared error and minimum mean squared error at $\theta_1 = \rho \frac{C_Y}{C_X}$ of $t_8$ is given by

$$B(t_8) \approx \frac{1}{n} \left[ \frac{1}{r} \left( \theta_1^2 C^2_X - \theta_1 \rho C_Y C_X \right) \right]$$

(6.8)

$$M(t_8) \approx \bar{Y}^2 \left[ \frac{1}{r} \left( \frac{1}{N} C^2_Y + \theta_1^2 \frac{1}{n} C^2_X - 2 \theta_1 \left( \frac{1}{r} \frac{1}{n} \right) \rho C_Y C_X \right) \right]$$

(6.9)

$$M(t_8)_{\text{min}} \approx \left( \frac{1}{r} \frac{n}{N} \right) \bar{S}^2_Y - \left( \frac{1}{r} \frac{n}{N} \right) \bar{S}^2_{XY}$$

(6.10)

$$y_{g_i} = \frac{1}{n-r} \left[ \frac{n y_r \bar{X}}{\theta_2 \bar{x} + (1 - \theta_2) \bar{x}} - r y_r \right] \quad \text{if } i \in R$$

(6.11)

Under this method, the point estimator of $\bar{Y}$ is

$$t_9 = \frac{\bar{y}_r \bar{X}}{\theta_2 \bar{x} + (1 - \theta_2) \bar{x}}$$

(6.12)

Lemma 3: The bias, mean squared error and minimum mean squared error at $\theta_2 = \rho \frac{C_Y}{C_X}$ of $t_9$ is given by

$$B(t_9) \approx \frac{1}{n} \left[ \frac{1}{r} \left( \theta_2^2 C^2_X - \theta_2 \rho C_Y C_X \right) \right]$$

(6.13)

$$M(t_9) \approx \bar{Y}^2 \left[ \frac{1}{r} \left( \frac{1}{N} C^2_Y + \theta_2^2 \frac{1}{n} C^2_X - 2 \theta_2 \left( \frac{1}{r} \frac{1}{n} \right) \rho C_Y C_X \right) \right]$$

(6.14)

$$M(t_9)_{\text{min}} \approx \left( \frac{1}{r} \frac{n}{N} \right) \bar{S}^2_Y - \left( \frac{1}{r} \frac{n}{N} \right) \bar{S}^2_{XY}$$

(6.15)

$$y_{10_i} = \frac{1}{n-r} \left[ \frac{n y_r \bar{X}}{\theta_3 \bar{x} + (1 - \theta_3) \bar{x}} - r y_r \right] \quad \text{if } i \in R$$

(6.16)

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

$$t_{10} = \frac{\bar{y}_r \bar{X}}{\theta_3 \bar{x} + (1 - \theta_3) \bar{x}}$$

(6.17)

Lemma 4: The bias, variance and minimum variance at $\theta_3 = \rho \frac{C_Y}{C_X}$ of $t_{10}$ is given by

$$B(t_{10}) \approx \frac{1}{r} \left( \frac{1}{N} \right) \bar{Y} \left( \theta_3^2 C^2_X - \theta_3 \rho C_Y C_X \right)$$

(6.18)

$$M(t_{10}) \approx \bar{Y}^2 \left[ \frac{1}{r} \left( \frac{1}{N} C^2_Y + \theta_3^2 C^2_X - 2 \theta_3 \rho C_Y C_X \right) \right]$$

(6.19)

$$M(t_{10})_{\text{min}} = \left( \frac{1}{r} \frac{n}{N} \right) \bar{S}^2_Y$$

(6.20)
Discussion

We considered, in the present research paper the study of some imputation methods in presence of missing observations under two phase sampling design while the number of responds is constant. But in practice it is not possible and the number of missing observations may be varying sample to sample. In such case the authors also extended suggested methods in case when number of respondent is varying. 

Conclusion

The proposed estimators are useful when some observations are missing in the sample and population mean of auxiliary information is unknown. Table-2 and 3, clearly indicates that the class of suggested estimators are more efficient in design I than design II. So, we can conclude that design I is better than design II. Table-4 shows bias and m.s.e for population A and B for Ahmed et al. (2006). It is also observed from table-5 that the suggested strategies are very close with Ahmed et al. 6.

References