Research Journal of Mathematical and Statistical Sciences ______________________________ISSN 2320–6047
Vol. **1(9),** 1-10, October **(2013)** *Res. J. Mathematical & Statistical Sci.*

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

Narendra Singh Thakur[1], Kalpana Yadav[2] and Sharad Pathak[3]
[1]Department of Mathematics and Statistics, Banasthali University, Banasthali, Rajasthan, INDIA
[2]Department of Statistics, Delhi University, Delhi, INDIA
[3]Department of Mathematics and Statistics, Dr. H.S. Gour Central Univesity, 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[1,2]. 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 secondary 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^c and ii. 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 i^{th} 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 $x_s = \{x_i : i \in S\}$ and for $i' \in S'$, the data $\{x_{i'} : i' \in S'\}$ are known with mean $\bar{x} = (n)^{-1}\sum_{i'=1}^{n'} x_{i'}$ and $\bar{x}' = (n')^{-1}\sum_{i'=1}^{n'} x_{i'}$ respectively[3]. The symbols that used are: $\bar{X}$, $\bar{Y}$: the population mean of x and y respectively; $\bar{x}$, $\bar{y}$: the sample mean of x and y respectively;
$\bar{x}_r$, $\bar{y}_r$: the sample mean of x and y respectively; ρ_{xy} : the correlation coefficient between x and y;
S_x^2, S_y^2 : the population mean squares of x and y respectively; C_x, C_y : the coefficient of variation of x and y respectively;

$$\delta_1 = \left(\frac{1}{r} - \frac{1}{n'}\right);\ \delta_2 = \left(\frac{1}{n} - \frac{1}{n'}\right);\ \delta_3 = \left(\frac{1}{n'} - \frac{1}{N}\right);\ \delta_4 = \left(\frac{1}{r} - \frac{1}{N-n'}\right);\ \delta_5 = \left(\frac{1}{n} - \frac{1}{N-n'}\right);\ f_1 = \frac{r}{n},$$

$$E = \frac{(\delta_{13} - \delta_4)(\delta_3 + \delta_5)}{\left[\delta_{13}(\delta_3 + \delta_5) - \{\ \delta_5^2 + (\delta_4 - \delta_5)(\delta_3 + \delta_5)\}\right]},\ F = \frac{(\delta_{14} - \delta_4)(\delta_3 + \delta_5)}{\left[\delta_{15}(\delta_3 + \delta_5) - \delta_5^2\right]},\ G = \frac{(\delta_{16} - \delta_4)(\delta_3 + \delta_4)}{\left[\delta_{16}(\delta_3 + \delta_4) - \delta_4^2\right]}.$$

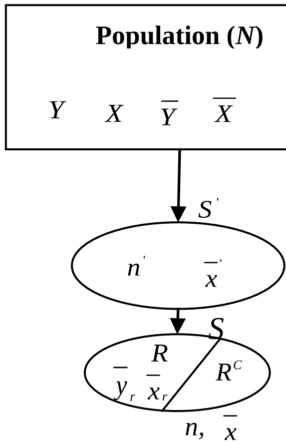

Figure-1
Sample S is as a sub-sample from sample S'

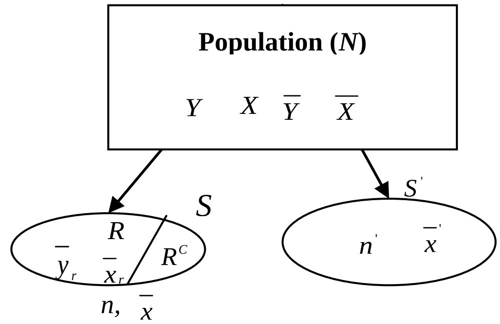

Figure-2
Sample S is independent to sample S' without replacing S' in the population

Large Sample Approximations

Let us consider $\bar{y}_r = \bar{Y}(1+e_1)$; $\bar{x}_r = \bar{X}(1+e_2)$; $\bar{x} = \bar{X}(1+e_3)$ and $\bar{x}' = \bar{X}(1+e_3')$. Now by using the concept of double sampling scheme and the mechanism of MCAR[4], for given *r*, *n* and n' we have:

Designs	$E(e_i)$	$E(e_3')$	$E(e_1^2)$	$E(e_2^2)$	$E(e_3^2)$	$E(e_3'^2)$
I	0	0	$\delta_1 C_Y^2$	$\delta_1 C_X^2$	$\delta_2 C_X^2$	$\delta_3 C_x^2$
II	0	0	$\delta_4 C_Y^2$	$\delta_4 C_X^2$	$\delta_5 C_X^2$	$\delta_3 C_x^2$

Designs	$E(e_1 e_2)$	$E(e_1 e_3)$	$E(e_1 e_3')$	$E(e_2 e_3)$	$E(e_2 e_3')$	$E(e_3 e_3')$
I	$\delta_1 \rho\, C_Y\, C_X$	$\delta_2 \rho\, C_Y C_X$	$\delta_3 \rho C_Y\, C_X$	$\delta_2\, C_X^2$	$\delta_3\, C_X^2$	$\delta_3\, C_X^2$
II	$\delta_4 \rho\, C_Y\, C_X$	$\delta_5 \rho\, C_Y C_X$	0	$\delta_5\, C_X^2$	0	0

Proposed Strategies

Let y'_{ji} 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:

$$y'_{7i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r + \dfrac{1}{(1-f_1)}\left[k_1\left(\bar{x}' - \bar{x}\right) + (1-f_1)k_2\left(x_i - \bar{x}_r\right)\right] & if \quad i \in R^C \end{cases} \quad \textbf{(3.1)}$$

under this strategy, the point estimator of $\bar{Y}$ is $t_7' = \bar{y}_r + k_1\left(\bar{x}' - \bar{x}\right) + k_2\left(\bar{x} - \bar{x}_r\right) \ldots (3.2)$

$$y'_{8i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{\bar{y}_r}{(1-f_1)}\left[\dfrac{\left(x_i(1-f_1) + f_1 \bar{x}_r\right)}{\theta_1 \bar{x}_r + (1-\theta_1)\bar{x}} - f_1\right] & if \quad i \in R^C \end{cases} \quad \textbf{(3.2)}$$

under this , the estimator of $\bar{Y}$ is

$$t_8' = \frac{\bar{y}_r \bar{x}}{\theta_1 \bar{x}_r + (1-\theta_1)\bar{x}} \quad \textbf{(3.3)}$$

$$y'_{9i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{\bar{y}_r}{(1-f_1)}\left[\dfrac{\bar{x}_r}{\theta_2 \bar{x} + (1-\theta_2)\bar{x}'} - f_1\right] & if \quad i \in R^C \end{cases} \quad \textbf{(3.4)}$$

Hence the estimator of $\overline{Y}$ is $$t_9^{'} = \frac{\overline{y}_r \overline{x}^{'}}{\theta_2 \overline{x} + (1-\theta_2)\overline{x}^{'}} \quad \textbf{(3.5)}$$

$$y_{10i}^{'} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{\overline{y}_r}{(1-f_1)}\left[\dfrac{\overline{x}^{'}}{\theta_3 \overline{x}_r + (1-\theta_3)\overline{x}^{'}} - f_1\right] & if \quad i \in R^C \end{cases} \quad \textbf{(3.6)}$$

Hence the estimator of $\overline{Y}$ is $$t_{10}^{'} = \frac{\overline{y}_r \overline{x}^{'}}{\theta_3 \overline{x}_r + (1-\theta_3)\overline{x}^{'}} \quad \textbf{(3.7)}$$

Bias and M.S.E. of Proposed Methods

Let $B(.)_t$ and $M(.)_t$ 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 $t_{10}^{'}$. The proofs of all these results are similar and therefore we will proof 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 $e_3^{'}$ could be expressed:

$$t_7^{'} = \overline{Y}(1+e_1) + k_1 \overline{X}\left(e_3^{'} - e_3\right) + k_2 \overline{X}\left(e_3 - e_2\right) \quad \textbf{(4.1)}$$

by ignoring the terms $E\left[e_i^r e_j^s\right]$, $E\left[e_i^r \left(e_j^{'}\right)^s\right]$ 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: $$t_7^{'} = \overline{y}_r + k_1\left(\overline{x}^{'} - \overline{x}\right) + k_2\left(\overline{x} - \overline{x}_r\right)$$
$$= \overline{Y}(1+e_1) + k_1 \overline{X}\left(e_3^{'} - e_3\right) + k_2 \overline{X}\left(e_3 - e_2\right)$$

The estimator $t_7^{'}$ is an unbiased estimator under both the designs *I* and *II* i.e.

$$B\left[t_7^{'}\right]_I = 0 \quad \textbf{(4.2)}$$
$$B\left[t_7^{'}\right]_{II} = 0 \quad \textbf{(4.3)}$$

Proof:
$$B\left(t_7^{'}\right)_I = E\left[t_7^{'} - \overline{Y}\right]_I = \overline{Y} - \overline{Y} = 0$$
$$B\left(t_7^{'}\right)_{II} = E\left[t_7^{'} - \overline{Y}\right]_{II} = \overline{Y} - \overline{Y} = 0$$

The variance of $t_7^{'}$, under design *I* and *II*, upto first order of approximation could be written as:

$$V\left(t_7^{'}\right)_I = \delta_1 S_Y^2 + (\delta_2 - \delta_3)\left(k_1^2 S_X^2 - 2k_1 \rho S_Y S_X\right) + (\delta_1 - \delta_2)\left(k_2^2 S_X^2 - 2k_2 \rho S_Y S_X\right) \quad \textbf{(4.4)}$$
$$V\left(t_7^{'}\right)_{II} = \delta_4 S_Y^2 + (\delta_3 + \delta_5) k_1^2 S_X^2 - 2k_1 \delta_5 \rho S_Y S_X + (\delta_4 - \delta_5)\left(k_2^2 S_X^2 - 2k_2 \rho S_Y S_X\right) \quad \textbf{(4.5)}$$

Proof: $$V\left(t_7^{'}\right) = E\left[t_7^{'} - \overline{Y}\right]^2 = E\left[\overline{Y} e_1 + k_1 \overline{X}\left(e_3^{'} - e_3\right) + k_2 \overline{X}\left(e_3 - e_2\right)\right]^2$$
$$= E\Big[\overline{Y}^2 e_1^2 + k_1^2 \overline{X}^2\left(e_3^{'} - e_3\right)^2 + k_2^2 \overline{X}^2\left(e_3 - e_2\right)^2 + 2k_1 \overline{Y}\overline{X}\left(e_3^{'} - e_3\right)e_1$$
$$+ 2k_1 k_2 \overline{X}^2\left(e_3^{'} - e_3\right)\left(e_3 - e_2\right) + 2k_2 \overline{Y}\overline{X}\left(e_3 - e_2\right)e_1\Big]$$
$$= E\Big[\overline{Y}^2 e_1^2 + k_1^2 \overline{X}^2\left(e_3^{'2} + e_3^2 - 2e_3 e_3^{'}\right) + k_2^2 \overline{X}^2\left(e_3^2 + e_2^2 - 2e_2 e_3\right) + 2k_1 \overline{Y}\overline{X}\left(e_1 e_3^{'} - e_1 e_3\right)$$
$$+ 2k_1 k_2 \overline{X}^2\left(e_3 e_3^{'} - e_3^2 - e_2 e_3^{'} + e_2 e_3\right) + 2k_2 \overline{Y}\overline{X}\left(e_1 e_3 - e_1 e_2\right)\Big]$$

Under Design *I* (Using (4.6))

$$V(t_7^{'})_I = \Big[\bar{Y}^2\ \delta_1 C_Y^2 + k_1^2 \bar{X}^2 \left(\delta_3 C_X^2 + \delta_2 C_X^2 - 2\delta_3 C_X^2\right) + k_2^2 \bar{X}^2 \left(\delta_2 C_X^2 + \delta_1 C_X^2 - 2\delta_2 C_X^2\right)$$

$$+ 2k_1 \bar{Y}\bar{X}\left(\delta_3 \rho C_Y C_X - \delta_2 \rho C_Y C_X\right) + k_1 k_2 \bar{X}^2\left(\delta_3 C_X^2 - \delta_2 C_X^2 - \delta_3 C_X^2 + \delta_2 C_X^2\right)$$

$$+ 2k_2 \bar{Y}\bar{X}\left(\delta_2 \rho C_Y C_X - \delta_1 \rho C_Y C_X\right)\Big]$$

$$= \Big[\bar{Y}^2\ \delta_1 C_Y^2 + k_1^2 \bar{X}^2 C_X^2 (\delta_2 - \delta_3) + + k_2^2 \bar{X}^2 C_X^2 (\delta_1 - \delta_2)$$

$$+ 2k_1 \bar{Y}\bar{X}(\delta_3 - \delta_2)\rho C_Y C_X + 2k_2 \bar{Y}\bar{X}(\delta_2 - \delta_1)\rho C_Y C_X\Big]$$

$$= \left[\delta_1 S_Y^2 + (\delta_2 - \delta_3)\left\{k_1^2 S_X^2 - 2k_1 \rho S_Y S_X\right\} + (\delta_1 - \delta_2)\left\{k_2^2 S_X^2 - 2k_2 \rho S_Y S_X\right\}\right]$$

Under Design *II* (Using (4.6))

$$V(t_7^{'})_{II} = \Big[\bar{Y}^2\ \delta_4 C_Y^2 + k_1^2 \bar{X}^2\left(\delta_3 C_X^2 + \delta_5 C_X^2\right) + k_2^2 \bar{X}^2\left(\delta_5 C_X^2 + \delta_4 C_X^2 - 2\delta_5 C_X^2\right)$$

$$+ 2k_1 \bar{Y}\bar{X}\left(-\delta_5 \rho C_Y C_X\right) + 2k_1 k_2 \bar{X}^2\left(-\delta_5 C_X^2 + \delta_5 C_X^2\right) + 2k_2 \bar{Y}\bar{X}\left(\delta_5 \rho C_Y C_X - \delta_4 \rho C_Y C_X\right)\Big]$$

$$= \Big[\bar{Y}^2\ \delta_4 C_Y^2 + k_1^2 S_X^2 (\delta_3 + \delta_5) + k_2^2 S_X^2 (\delta_4 - \delta_5) - 2k_1 \delta_5 \rho S_Y S_X - 2k_2 (\delta_4 - \delta_5)\rho S_Y S_X\Big]$$

$$= \delta_4 S_Y^2 + (\delta_3 + \delta_5) k_1^2 S_X^2 - 2k_1 \delta_5 \rho S_Y S_X + (\delta_4 - \delta_5)\left(k_2^2 S_X^2 - 2k_2 \rho S_Y S_X\right)$$

The minimum variance of the $t_7^{'}$ is

$$\left[V(t_7^{'})_I\right]_{Min} = \left[\delta_1 - (\delta_1 - \delta_3)\rho^2\right] S_Y^2 \quad \textbf{(4.7)}$$

$$\left[V(t_7^{'})_{II}\right]_{Min} = \left[\delta_4 - \left(\delta_3 \delta_4 + \delta_4 \delta_5 - \delta_3 \delta_5\right)(\delta_3 + \delta_5)^{-1} \rho^2\right] S_Y^2 \quad \textbf{(4.8)}$$

Proof:
First differentiate (4.4) with respect to k_1 and k_2 and then equate to zero, we get

$$\frac{d}{dk_1}\left[V(t_7^{'})_I\right] = 0 \Rightarrow k_1 = \rho \frac{S_Y}{S_X} \text{ and } \frac{d}{dk_2}\left[V(t_7^{'})_I\right] = 0 \Rightarrow k_2 = \rho \frac{S_Y}{S_X}$$

After replacing value of β_1 in (4.4) , we obtained

$$\left[V(t_7^{'})_I\right]_{Min} = \left[\delta_1 - (\delta_1 - \delta_3)\rho^2\right] S_Y^2$$

Similar to (i), we proceed for (4.5), we have

$$\frac{d}{dk_1}\left[V(t_7^{'})_{II}\right] = 0 \Rightarrow k_1 = \left(\frac{\delta_5}{\delta_3 + \delta_5}\right)\rho \frac{S_Y}{S_X} \text{ and } \frac{d}{dk_2}\left[V(t_7^{'})_{II}\right] = 0 \Rightarrow k_2 = \rho \frac{S_Y}{S_X}$$

$$\left[V(t_7^{'})_{II}\right]_{Min} = \left[\delta_4 - \left(\delta_3 \delta_4 + \delta_4 \delta_5 - \delta_3 \delta_5\right)(\delta_3 + \delta_5)^{-1} \rho^2\right] S_Y^2$$

Theorem 4.2: The estimator $t_8^{'}$ in terms of e_1, e_2, e_3 and $e_3^{'}$ is

$$t_8^{'} = \bar{Y}\left[1 + e_1 + \theta_1\left(e_3 - e_2 - e_1 e_2 + e_1 e_3 + (1 - 2\theta_1)e_2 e_3 + \theta_1 e_2^2 - (1 - \theta_1)e_3^2\right)\right] \quad \textbf{(4.9)}$$

The bias of the estimator $t_8^{'}$ under design *I* and *II* respectively is

$$B\left(t_8^{'}\right)_I = \bar{Y}(\delta_1 - \delta_2)\left(\theta_1^2 C_X^2 - \theta_1 \rho C_Y C_X\right) \quad \textbf{(4.10)}$$

$$B\left(t_8^{'}\right)_{II} = \bar{Y}(\delta_4 - \delta_5)\left(\theta_1^2 C_X^2 - \theta_1 \rho C_Y C_X\right) \quad \textbf{(4.11)}$$

Mean squared error of $t_8^{'}$ under design under design *I* and *II* respectively is:

$$M\left(t_8^{'}\right)_I = \bar{Y}^2\left[\delta_1 C_Y^2 + (\delta_1 - \delta_2)\left(\theta_1^2 C_X^2 - 2\theta_1 \rho C_Y C_X\right)\right] \quad \textbf{(4.12)}$$

$$M\left(t_8^{'}\right)_{II} = \bar{Y}^2\left[\delta_4 C_Y^2 + (\delta_4 - \delta_5)\left\{\theta_1^2 C_X^2 - 2\theta_1 \rho C_Y C_X\right\}\right] \quad \textbf{(4.13)}$$

The minimum m.s.e. of $t_8^{'}$ is

$$\left[M\left(t_8^{'}\right)_I\right]_{Min} = \left[\delta_1 - (\delta_1 - \delta_2)\rho^2\right]S_Y^2 \quad \text{when } \theta_1 = \rho\frac{C_Y}{C_X} \quad \textbf{(4.14)}$$

$$\left[M\left(t_8^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - (\delta_4 - \delta_5)\rho^2\right]S_Y^2 \quad \text{when } \theta_1 = \rho\frac{C_Y}{C_X} \quad \textbf{(4.15)}$$

Theorem 4.3:
The estimator $t_9^{'}$ in terms of e_1, e_2, e_3 and $e_3^{'}$ is

$$t_9^{'} = \bar{Y}\left[1 + e_1 + \theta_2\left(e_3^{'} - e_3 + e_1 e_3^{'} - e_1 e_3 - (1 + 2\theta_2)e_3 e_3^{'} + \theta_2 e_3^2 + (1 + \theta_2)e_3^{'2}\right)\right] \quad \textbf{(4.16)}$$

The bias of the estimator $t_9^{'}$ under design *I* and *II*respectively is:

$$B\left(t_9^{'}\right)_I = \bar{Y}(\delta_2 - \delta_3)\left(\theta_2^2 C_X^2 - \theta_2 \rho C_Y C_X\right) \quad \textbf{(4.17)}$$

$$B\left(t_9^{'}\right)_{II} = \bar{Y}\left(\left[\theta_2^2(\delta_3 + \delta_5) + \delta_3\theta_2\right]C_X^2 - \theta_2 \delta_5 \rho C_Y C_X\right) \quad \textbf{(4.18)}$$

Mean squared error of $t_9^{'}$ under design *I* and *II*respectively is:

$$M\left(t_9^{'}\right)_I = \bar{Y}^2\left[\delta_1 C_Y^2 + (\delta_2 - \delta_3)\left(\theta_2^2 C_X^2 - 2\theta_2 \rho C_Y C_X\right)\right] \quad \textbf{(4.19)}$$

$$M\left(t_9^{'}\right)_{II} = \bar{Y}^2\left[\delta_4 C_Y^2 + (\delta_3 + \delta_5)\theta_2^2 C_X^2 - 2\theta_2 \delta_5 \rho C_Y C_X\right] \quad \textbf{(4.20)}$$

The minimum m.s.e. of $t_9^{'}$ is

$$\left[M\left(t_9^{'}\right)_I\right]_{Min} = \left[\delta_1 - (\delta_2 - \delta_3)\rho^2\right]S_Y^2 \quad \text{when } \theta_2 = \rho\frac{C_Y}{C_X} \quad \textbf{(4.21)}$$

$$\left[M\left(t_9^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - \delta_5^2(\delta_3 + \delta_5)^{-1}\rho^2\right]S_Y^2 \text{ when } \theta_2 = \left(\frac{\delta_5}{\delta_3 + \delta_5}\right)\rho\frac{C_Y}{C_X} \quad \textbf{(4.22)}$$

Theorem 4.4: The estimator $t_{10}^{'}$ in terms of e_1, e_2, e_3 and $e_3^{'}$ is

$$t_{10}^{'} = \bar{Y}\left[1 + e_1 + \theta_3\left(e_3^{'} - e_2 + e_1 e_3^{'} - e_1 e_2 + \theta_3 e_2^2 + \theta_3 e_3^{'2} - e_3^{'2} - e_2 e_3^{'}\right)\right]$$

The bias of the estimator $t_{10}^{'}$ under design *I* and *II*respectively is:

$$B\left(t_{10}^{'}\right)_I = \bar{Y}\left(\theta_3^2(\delta_1 + \delta_3)C_X^2 - 2\delta_3\theta_3 C_X^2 - \theta_3(\delta_1 - \delta_3)\rho C_Y C_X\right) \quad \textbf{(4.23)}$$

$$B\left(t_{10}^{'}\right)_{II} = \bar{Y}\left(\theta_3^2(\delta_4 + \delta_3)C_X^2 - \delta_3\theta_3 C_X^2 - \theta_3 \delta_4 \rho C_Y C_X\right) \quad \textbf{(4.24)}$$

Mean squared error of underdesign *I* and *II*respectively is:

$$M\left(t_{10}^{'}\right)_I = \overline{Y}^2\left[\delta_1 C_Y^2 + (\delta_1 - \delta_3)\left(\theta_3^2 C_X^2 - 2\theta_3 \rho C_Y C_X\right)\right] \quad \textbf{(4.25)}$$

$$M\left(t_{10}^{'}\right)_{II} = \overline{Y}^2\left[\delta_4 C_Y^2 + (\delta_3 + \delta_4)\theta_3^2 C_X^2 - 2\theta_3 \delta_4 \rho C_Y C_X\right] \quad \textbf{(4.26)}$$

The minimum m.s.e. of $t_{10}^{'}$ is

$$\left[M\left(t_{10}^{'}\right)_I\right]_{Min} = \left[\delta_1 - (\delta_1 - \delta_3)\rho^2\right] S_Y^2 \quad \text{when } \theta_3 = \rho\frac{C_Y}{C_X} \quad \textbf{(4.27)}$$

$$\left[M\left(t_{10}^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - \delta_4^2(\delta_3 + \delta_4)^{-1}\rho^2\right] S_Y^2 \text{ when } \theta_3 = \left(\frac{\delta_4}{\delta_3 + \delta_4}\right)\rho\frac{C_Y}{C_X} \quad \textbf{(4.28)}$$

Comparisons

$$\Delta_{13} = \min\left[V(t_7)\right] - \min\left[V(t_7^{'})_I\right]$$

$$= \left[\frac{1}{n^{'}} - \frac{1}{N}\right] S_Y^2 + \left[\frac{2}{N} - \frac{2}{n^{'}}\right]\rho^2 S_y^2$$

$\left(t_7^{'}\right)_{\text{I}}$ is better than t_7, if $\Delta_{13} > 0$

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

$$\Rightarrow -\frac{1}{2} < \rho < \frac{1}{2}$$

$$\Delta_{14} = \min\left[V(t_7)\right] - \min\left[V(t_7^{'})_{II}\right]$$

$$= \left[\delta_{13} - \delta_4\right] S_Y^2 - \left[\left\{(\delta_3 + \delta_5)^{-1}\delta_5^2 + (\delta_4 - \delta_5)\right\} - \delta_{13}\right]\rho^2 S_Y^2$$

$\left(t_7^{'}\right)_{\text{II}}$ is better than t_7, if $\Delta_{14} > 0$

$$\Rightarrow \rho^2 < \frac{(\delta_{13} - \delta_4)(\delta_3 + \delta_5)}{\left[\delta_{13}(\delta_3 + \delta_5) - \left\{\delta_5^2 + (\delta_4 - \delta_5)(\delta_3 + \delta_5)\right\}\right]}$$

$$\Rightarrow -E < \rho < E$$

$$\Delta_{15} = \min\left[V(t_8)\right] - \min\left[V(t_8^{'})_I\right] = \left[\frac{1}{n^{'}} - \frac{1}{N}\right] S_Y^2$$

$\left(t_8^{'}\right)_{\text{I}}$ is better than t_8, if $\Delta_{15} > 0$

$$\Rightarrow \left[\frac{N - n^{'}}{n^{'}N}\right] > 0 \Rightarrow N - n^{'} > 0 \quad \Rightarrow n^{'} < N$$

which is always true.

$$\Delta_{16} = \min\left[V(t_8)\right] - \min\left[V(t_8^{'})_{II}\right] = \left[\frac{1}{N - n^{'}} - \frac{1}{N}\right] S_Y^2$$

$\left(t_8^{'}\right)_{\text{II}}$ is better than t_8, if $\Delta_{16} > 0$

$$\Rightarrow \left[\frac{N - N + n^{'}}{N\left(N - n^{'}\right)}\right] > 0 \qquad \Rightarrow n^{'} > 0$$

$$\Delta_{17} = \min\left[V(t_9)\right] - \min\left[V(t_9^{'})_I\right]$$

$$= \left[\frac{1}{n^{'}} - \frac{1}{N}\right] S_Y^2 + \left[\frac{2}{N} - \frac{2}{n^{'}}\right]\rho^2 S_y^2$$

$(t'_{17})_{\text{I}}$ is better than t_9, if $\Delta_{17} > 0 \quad \Rightarrow \rho^2 < \frac{1}{2} \quad \Rightarrow -\frac{1}{2} < \rho < \frac{1}{2}$

$\Delta_{18} = \min[V(t_9)] - \min[V(t'_9)_{II}] = [\delta_{14} - \delta_4]\, S_Y^2 - [\delta_{15} - (\delta_3 + \delta_5)^{-1}\delta_5^2]\, \rho^2 S_Y^2 \quad (t'_9)_{\text{II}}$ is better than t_9, if $\Delta_{18} > 0$

$$\Rightarrow \rho^2 < \frac{(\delta_{14} - \delta_4)(\delta_3 + \delta_5)}{[\delta_{15}(\delta_3 + \delta_5) - \delta_5^2]} \quad \Rightarrow -F < \rho < F$$

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

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

which is always true.

$\Delta_{20} = \min[V(t_{10})] - \min[V(t'_{10})_{II}] = [\delta_{16} - \delta_4]\, S_Y^2 - [\delta_{16} - (\delta_3 + \delta_4)^{-1}\delta_4^2]\, \rho^2 S_Y^2$

$(t'_{10})_{\text{II}}$ is better than t_{10}, if

$$\Delta_{20} > 0 \quad \Rightarrow \rho^2 < \frac{(\delta_{16} - \delta_4)(\delta_3 + \delta_4)}{[\delta_{16}(\delta_3 + \delta_4) - \delta_4^2]} \quad \Rightarrow -G < \rho < G$$

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-1
Population Parameters

Population	N	$\bar{Y}$	$\bar{X}$	S_Y^2	S_X^2	ρ	C_X	C_Y
A	200	42.485	18.515	199.0598	48.5375	0.8652	0.3763	0.3321
B	8306	253.75	343.316	338006	862017	0.522231	2.70436	2.29116

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-2
Bias and MSE for Population – A

Estimators	DESIGN *I*		DESIGN *II*	
	Bias	**MSE**	**Bias**	**MSE**
t'_7	0	10.91418	0	38.71673
t'_8	-1.40126E-06	10.41748	-5.95325E-05	12.31328
t'_9	2.66906E-08	35.33217	.26202	36.78069
t'_{10}	-.025405	9.255346	1325.124	11.29167

Table-3
Bias and MSE for Population – B

Estimators	DESIGN *I*		DESIGN *II*	
	Bias	**MSE**	**Bias**	**MSE**
t'_7	0	16300.3	0	22485.14
t'_8	0.00000381	16403.58	0.00000974	16518.98
t'_9	0.00000006	21754.44	-0.26502	22339.4
t'_{10}	-0.34747	15793.29	9.819971	16384.03

Table-4
Bias and MSE for Population A and B for Ahmed et al. (2006)

Estimators	**Population A**		**Population B**	
	Bias	**MSE**	**Bias**	**MSE**
t_7	0	9.759633	0	16358.62
t_8	-.0000595	12.73984	-0.09258	16531.89
t_9	-.0000068	35.83645	-0.09527097	22319.77
t_{10}	-.0000663	9.759633	0.095271	16358.62

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

$$E_i = \frac{Opt\left[M\left(t_i^{'}\right)_j\right]}{Opt\left[M\left(t_i\right)\right]}; \qquad i = 7,8,9,10; \qquad j = I, II \qquad \textbf{…(*)}$$

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

Table-5
Efficiency for Population A and B over Ahmed et al. (2006)[6]

Estimators	**Population A**		**Population B**	
	Design *I*	**Design *II***	**Design *I***	**Design *II***
E_7	1.118298	3.967027	0.996435	1.374513
E_8	0.817709	0.966518	0.992239	0.999219
E_9	0.985928	1.026349	0.974671	1.000879
E_{10}	0.948329	1.156977	0.965441	1.001553

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[6]. Let y_{ji} 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 :

$$y_{7i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r + \dfrac{nk_1}{(n-r)}\left(\bar{X} - \bar{x}\right) + k_2\left(x_i - \bar{x}_r\right) & if \quad i \in R^C \end{cases} \qquad \textbf{(6.1)}$$

Under this method, the point estimator of $\bar{Y}$ is $\quad t_7 = \bar{y}_r + k_1\left(\bar{X} - \bar{x}\right) + k_2\left(\bar{x} - \bar{x}_r\right)$ **(6.2)**

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

$$B[t_7] = 0 \qquad \textbf{(6.3)}$$

$$V(t_7) = \left(\frac{1}{r} - \frac{1}{N}\right)S_Y^2 - 2S_{XY}\left[k_1\left(\frac{1}{n} - \frac{1}{N}\right) + k_2\left(\frac{1}{r} - \frac{1}{n}\right)\right] + S_X^2\left[k_1^2\left(\frac{1}{n} - \frac{1}{N}\right) + k_2^2\left(\frac{1}{r} - \frac{1}{n}\right)\right] \qquad \textbf{(6.4)}$$

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

$$y_{8i} = \begin{cases} y_i & if \quad i \in R \\ \left[\dfrac{\overline{y_r}\left(x_i + \dfrac{r}{n-r}\bar{x}_r\right)}{\theta_1\bar{x}_r + \left(1-\theta_1\right)\bar{x}} - \dfrac{r}{n-r}\overline{y_r}\right] & if \quad i \in R^C \end{cases} \qquad \textbf{(6.6)}$$

Under this method, the point estimator of $\bar{Y}$ is $$t_8 = \frac{\bar{y}_r \bar{x}}{\theta_1 \bar{x}_r + (1-\theta_1)\bar{x}} \quad \textbf{(6.7)}$$

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

$$B(t_8) \approx \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y} \left(\theta_1^2 C_X^2 - \theta_1 \rho C_Y C_X\right) \quad \textbf{(6.8)}$$

$$M(t_8) \approx \bar{Y}^2 \left[\left(\frac{1}{r} - \frac{1}{N}\right) C_Y^2 + \theta_1^2 \left(\frac{1}{r} - \frac{1}{n}\right) C_X^2 - 2\theta_1 \left(\frac{1}{r} - \frac{1}{n}\right) \rho C_Y C_X\right] \quad \textbf{(6.9)}$$

$$M(t_8)_{\min} \approx \left(\frac{1}{r} - \frac{1}{N}\right) S_Y^2 - \left(\frac{1}{r} - \frac{1}{n}\right) \frac{S_{XY}^2}{S_X^2} \quad \textbf{(6.10)}$$

$$y_{9i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[\dfrac{n\bar{y}_r \bar{X}}{\theta_2 \bar{x} + (1-\theta_2)\bar{X}} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \quad \textbf{(6.11)}$$

Under this method, the point estimator of $\bar{Y}$ is $$t_9 = \frac{\bar{y}_r \bar{X}}{\theta_2 \bar{x} + (1-\theta_2)\bar{X}} \quad \textbf{(6.12)}$$

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

$$B(t_9) \approx \left(\frac{1}{n} - \frac{1}{N}\right) \bar{Y} \left(\theta_2^2 C_X^2 - \theta_2 \rho C_Y C_X\right) \quad \textbf{(6.13)}$$

$$M(t_9) \approx \bar{Y}^2 \left[\left(\frac{1}{r} - \frac{1}{N}\right) C_Y^2 + \theta_2^2 \left(\frac{1}{n} - \frac{1}{N}\right) C_X^2 - 2\theta_2 \left(\frac{1}{n} - \frac{1}{N}\right) \rho C_Y C_X\right] \quad \textbf{(6.14)}$$

$$M(t_9)_{\min} \approx \left(\frac{1}{r} - \frac{1}{N}\right) S_Y^2 - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{XY}^2}{S_X^2} \quad \textbf{(6.15)}$$

$$y_{10i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[\dfrac{n\bar{y}_r \bar{X}}{\theta_3 \bar{x}_r + (1-\theta_3)\bar{X}} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \quad \textbf{(6.16)}$$

Under this, the point estimator of population mean $\bar{Y}$ is $$t_{10} = \frac{\bar{y}_r \bar{X}}{\theta_3 \bar{x}_r + (1-\theta_3)\bar{X}} \quad \textbf{(6.17)}$$

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

$$B(t_{10}) \approx \left(\frac{1}{r} - \frac{1}{N}\right) \bar{Y} \left(\theta_3^2 C_X^2 - \theta_3 \rho C_Y C_X\right) \quad \textbf{(6.18)}$$

$$M(t_{10}) \approx \bar{Y}^2 \left(\frac{1}{r} - \frac{1}{N}\right) \left[C_Y^2 + \theta_3^2 C_X^2 - 2\theta_3 \rho C_Y C_X\right] \quad \textbf{(6.19)}$$

$$M(t_{10})_{\min} = \left(\frac{1}{r} - \frac{1}{N}\right) S_Y^2 \left(1 - \rho^2\right) \quad \textbf{(6.20)}$$

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. [8,9,10]

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 estimatorsare 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

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

2. Thakur N.S., Yadav K. and Pathak S., Estimation of mean in presence of missing data under two-phase sampling scheme, *Journal of Reliability and Statistical Studies*, **4(2)**, 93-104 **(2011)**

3. Thakur N.S., Yadav K. and Pathak S., Some imputation methods in double sampling scheme for estimation of population mean, *International Journal of Modern Engineering Research*, **2(1),** 200-207 **(2012)**

4. Rao J.N.K. and Sitter R.R., Variance estimation under two-phase sampling with application to imputation for missing data, *Biometrica*, **82,** 453-460 **(1995)**

5. Shukla D. and Thakur N.S., Estimation of mean with imputation of missing data using factor-type estimator, Statistics in Transition, **9(1)**, 33-48 **(2008)**

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

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

8. Thakur N.S., Yadav K. and Pathak S., Imputation using regression estimators for estimating population mean in two-phase sampling, *Journal of Reliability and Statistical Studies*, **5(2),** 21-31 **(2012)**

9. Thakur N.S., Yadav K. and Pathak S., Mean estimation with imputation in two-phase sampling, *International Journal of Modern Engineering Research*, **2(5),** 3561-3571 **(2012)**

10. Thakur N.S., Yadav K. and Pathak S., On mean estimation with imputation in two-phase sampling design, *Research Journal of Mathematical and Statistical Sciences*, **1(3),** 1-9 **(2013)**