Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 4, Issue 2 (2011): 93-104

ESTIMATION OF MEAN IN PRESENCE OF MISSING DATA UNDER TWO-PHASE SAMPLING SCHEME

Narendra Singh Thakur[1], Kalpana Yadav[2] and Sharad Pathak[3]
[1,2] Centre for Mathematical Sciences (CMS), Banasthali University, Rajasthan
E Mail: nst_stats@yahoo.co.in, kalpana22yadav@gmail.com
[3]Department of Mathematics and Statistics, Dr. H. S. Gour Central Univesity, Sagar (M.P.)

Abstract

To estimate the population mean with imputation i.e. the technique of substituting missing data, there are a number of techniques available in literature like Ratio method of imputation, Compromised method of imputation, Mean method of imputation, Ahmed method of imputation, F-T method of imputation, and so on. If population mean of auxiliary information is unknown then these methods are not useful and the two-phase sampling is used to obtain the population mean. This paper presents some imputation methods of for missing values in two-phase sampling. Two different sampling designs in two-phase sampling are compared under imputed data. The bias and m.s.e of suggested estimators are derived in the form of population parameters using the concept of large sample approximation. Numerical study is performed over two populations using the expressions of bias and m.s.e and efficiency compared with Ahmed estimators.

Keywords: Estimation, Missing data, Bias, Mean squared error (M.S.E), Two-phase sampling, SRSWOR, Large sample approximations.

1. Introduction

To overcome the problem of missing observations or non-response in sample surveys, the technique of imputation is frequently used to replace the missing data. To deal with missing values effectively Kalton et al. (1981) and Sande (1979) suggested imputation that make an incomplete data set structurally complete and its analysis simple. Imputation may also be carried out with the aid of an auxiliary variate if it is available. For example Lee et al. (1994, 1995) used the information on an auxiliary variate for the purpose of imputation. Later Singh and Horn (2000) suggested a compromised method of imputation. Ahmed et al. (2006) suggested several new imputation based estimators that use the information on an auxiliary variate and compared their performances with the mean method of imputation. Shukla (2002) disussed F-T estimator under two-phase sampling and Shukla and Thakur (2008) have proposed estimation of mean with imputation of missing data using F-T estimator. Shukla et al. (2009) have discussed on utilization of non-response auxiliary population mean in imputation for missing observations and Shukla et al. (2009a) have discussed on estimation of mean under imputation of missing data using factor type estimator in two-phase sampling. Shukla et al. (2011) suggested linear combination based imputation method for missing data in sample. The objective of the present research work is to derive some imputation method for mean estimation in case population parameter of auxiliary information is unknown.

2. Notations

Let U = (U_1, U_2, U_3,…,U_N) be the finite population of size N and the character under study be denoted by Y and X be the auxiliary variable correlated with Y. 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 ($n < n'$) drawn in either two ways: One is as a sub-sample from sample S' (denoted by design *I*) as in fig. 1 and other is independent to sample S' (denoted by design *II*) as in fig. 2 without replacing S'. The sample S can be divided into two non-overlapping sub groups, the set of responding units, by R, and that of non- responding units by R^c and the number of responding units out of sampled n units be denoted b r $(r < n)$ For every unit $i \in R$ the value 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. The following symbols are used hereafter:

$\bar{X}$, $\bar{Y}$: 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;

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

$$\delta_5 = \left(\frac{1}{n} - \frac{1}{N-n'}\right);\ f_1 = \frac{r}{n},\ A = \frac{(\delta_6 - \delta_4)(\delta_3 + \delta_5)}{\left[\delta_7(\delta_3 + \delta_5) - \delta_5^2\right]};\ B = \frac{(\delta_8 - \delta_4)(\delta_3 + \delta_4)}{\left[\delta_8(\delta_3 + \delta_4) - \delta_4^2\right]}.$$

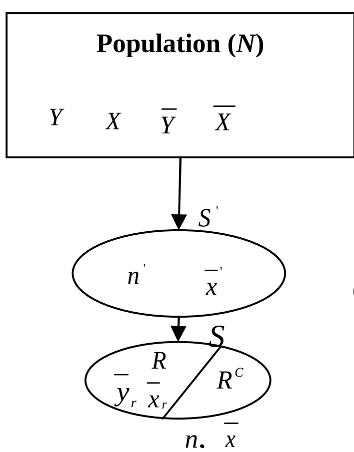

Fig. 1 [Design *I*]

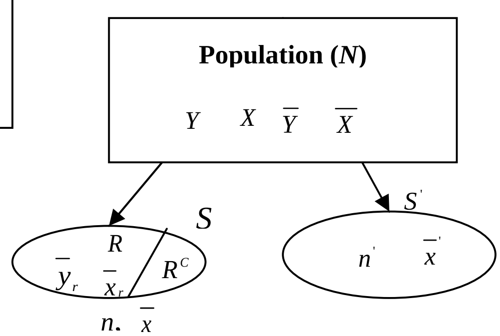

Fig. 2 [Design *II*]

3. Large Sample Approximations

Let $\bar{y}_r = \bar{Y}(1+e_1)$; $\bar{x}_r = \bar{X}(1+e_2)$; $\bar{x} = \bar{X}(1+e_3)$ and $\bar{x}' = \bar{X}(1+e_3')$, which implies the results $e_1 = \frac{\bar{y}_r}{\bar{Y}} - 1$; $e_2 = \frac{\bar{x}_r}{\bar{X}} - 1$; $e_3 = \frac{\bar{x}}{\bar{X}} - 1$ and $e_3' = \frac{\bar{x}'}{\bar{X}} - 1$. Now by using the concept of two-phase sampling and the the mechanism of MCAR, for given *r*, *n* and n' (see Rao and Sitter (1995)) 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

4. Proposed Strategies

Let y_{ji}' denotes the i^{th} observation of the j^{th} suggested imputation strategy and β_1, β_2, β_3 are constants such that the variance of obtained estimators of $\bar{Y}$ is minimum. We suggest the following tools of imputation:

(1)
$$y_{1i}' = \begin{cases} y_i & \text{if} \quad i \in R \\ \dfrac{\bar{y}_r}{(1-f_1)}\left[\left(\dfrac{\bar{x}'}{\bar{x}}\right)^{\beta_1} - f_1\right] & \text{if} \quad i \in R^C \end{cases} \tag{4.1}$$

under this strategy, the point estimator of $\bar{Y}$ is
$$t_1' = \bar{y}_r \left(\frac{\bar{x}'}{\bar{x}}\right)^{\beta_1} \tag{4.2}$$

(2)
$$y_{2i}' = \begin{cases} y_i & \text{if} \quad i \in R \\ \dfrac{\bar{y}_r}{(1-f_1)}\left[\left(\dfrac{\bar{x}}{\bar{x}_r}\right)^{\beta_2} - f_1\right] & \text{if} \quad i \in R^C \end{cases} \tag{4.3}$$

under this , the estimator of $\bar{Y}$ is
$$t_2' = \bar{y}_r \left(\frac{\bar{x}}{\bar{x}_r}\right)^{\beta_2} \tag{4.4}$$

(3)
$$y_{3i}' = \begin{cases} y_i & \text{if} \quad i \in R \\ \dfrac{\bar{y}_r}{(1-f_1)}\left[\left(\dfrac{\bar{x}'}{\bar{x}_r}\right)^{\beta_3} - f_1\right] & \text{if} \quad i \in R^C \end{cases} \tag{4.5}$$

Hence the estimator of $\overline{Y}$ is
$$t_3^{'} = \overline{y}_r \left(\frac{\overline{x}^{'}}{\overline{x}_r} \right)^{\beta_3} \quad (4.6)$$

Note: At $\beta_3 = 1\ (-1)$ then the estimator $t_3^{'}$ convert into ratio (product) type estimator in two-phase sampling scheme.

5. Properties of Proposed Estimators

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*. The properties of $t_1^{'}$, $t_2^{'}$ and $t_3^{'}$ are derived in the following theorems respectively. The proofs of all these results are similar and therefore we will proof only one of them i.e. theorem 5.1.

Theorem 5.1

(1) Estimator $t_1^{'}$ in terms of e_i; $i = 1,2,3$ and $e_3^{'}$ could be expressed:

$$t_1^{'} = \overline{Y}\left[1 + e_1 + \beta_1\left\{e_3^{'} - e_3 - e_1e_3 + e_1e_3^{'} - \beta_1 e_3 e_3^{'} + \frac{\beta_1+1}{2}e_3^2 + \frac{\beta_1-1}{2}e_3^{'2}\right\}\right] \quad (5.1)$$

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 $t_1^{'} = \overline{y}_r \left(\frac{\overline{x}^{'}}{\overline{x}}\right)^{\beta_1} = \overline{Y}(1+e_1)(1+e_3^{'})^{\beta_1}(1+e_3)^{-\beta_1}$

$$= \overline{Y}(1+e_1)\left(1 + \beta_1 e_3^{'} + \frac{\beta_1(\beta_1-1)}{2}e_3^{'2}\right)\left(1 - \beta_1 e_3 + \frac{\beta_1(\beta_1+1)}{2}e_3^2\right)$$

$$= \overline{Y}\left[1 + e_1 + \beta_1\left\{e_3^{'} - e_3 - e_1e_3 + e_1e_3^{'} - \beta_1 e_3 e_3^{'} + \frac{\beta_1+1}{2}e_3^2 + \frac{\beta_1-1}{2}e_3^{'2}\right\}\right]$$

(2) Bias of $t_1^{'}$ under design *I* and *II* is:

(i) $$B\left(t_1^{'}\right)_I = \overline{Y}\beta_1(\delta_2 - \delta_3)\left(\frac{\beta_1+1}{2}C_X^2 - \rho C_Y C_X\right) \quad (5.2)$$

(ii) $$B\left(t_1^{'}\right)_{II} = \overline{Y}\beta_1\left[\frac{1}{2}\{\beta_1(\delta_3+\delta_5) - (\delta_3-\delta_5)\}C_X^2 - \delta_5\rho C_Y C_X\right] \quad (5.3)$$

Proof **(i)** $B\left(t_1^{'}\right)_I = E\left[t_1^{'} - \overline{Y}\right]_I$

$$= \overline{Y}\,E\left[1 + e_1 + \beta_1\left\{e_3^{'} - e_3 - e_1e_3 + e_1e_3^{'} - \beta_1 e_3 e_3^{'} + \frac{\beta_1+1}{2}e_3^2 + \frac{\beta_1-1}{2}e_3^{'2}\right\} - 1\right]$$

$$= \overline{Y}\beta_1(\delta_2 - \delta_3)\left(\frac{\beta_1+1}{2}C_X^2 - \rho C_Y C_X\right)$$

(ii) $B\left(t_1^{'}\right)_{II} = E\left[t_1^{'} - \overline{Y}\right]_{II} = \overline{Y}\beta_1\left[\frac{1}{2}\{\beta_1(\delta_3+\delta_5) - (\delta_3-\delta_5)\}C_X^2 - \delta_5\rho C_Y C_X\right]$

(3) Mean squared error of t_1' under design *I* and *II*, upto first order of approximation could be written as:

(i) $$M(t_1')_I = \bar{Y}^2\left[\delta_1 C_Y^2 + (\delta_2 - \delta_3)(\beta_1^2 C_X^2 - 2\beta_1 \rho C_Y C_X)\right] \quad (5.4)$$

(ii) $$M(t_1')_{II} = \bar{Y}^2\left[\delta_4 C_Y^2 + (\delta_3 + \delta_5)\beta_1^2 C_X^2 - 2\delta_5 \beta_1 \rho C_Y C_X\right] \quad (5.5)$$

Proof $$M(t_1') = E\left[t_1' - \bar{Y}\right]^2 = \bar{Y}^2 E\left[1 + e_1 + \beta_1(e_3' - e_3) - 1\right]^2$$

$$= \bar{Y}^2 E\left[e_1^2 + \beta_1^2\left(e_3'^2 + e_3^2 - 2e_3 e_3'\right) + 2\beta_1\left(e_1 e_3' - e_1 e_3\right)\right] \quad (5.6)$$

(i) Under Design *I* (Using (5.6))

$$M(t_1')_I = \bar{Y}^2\left[\delta_1 C_Y^2 + (\delta_2 - \delta_3)(\beta_1^2 C_X^2 - 2\beta_1 \rho C_Y C_X)\right]$$

(ii) Under Design *II* (Using (5.6))

$$M(t_1')_{II} = \bar{Y}^2\left[\delta_4 C_Y^2 + (\delta_3 + \delta_5)\beta_1^2 C_X^2 - 2\delta_5 \beta_1 \rho C_Y C_X\right]$$

(4) Minimum mean squared error of t_1' is :

(i) $$\left[M(t_1')_I\right]_{Min} = \left[\delta_1 - (\delta_2 - \delta_3)\rho^2\right] S_Y^2 \quad \text{when } \beta_1 = \rho\frac{C_Y}{C_X} \quad (5.7)$$

(ii) $$\left[M(t_1')_{II}\right]_{Min} = \left[\delta_4 - (\delta_3 + \delta_5)^{-1}\delta_5^2\rho^2\right] S_Y^2 \text{ when } \beta_1 = \delta_5(\delta_3 + \delta_5)^{-1}\rho\frac{C_Y}{C_X} \quad (5.8)$$

Proof **(i)** First differentiate (5.4) with respect to β_1 and then equate to zero, we get

$$\frac{d}{d\beta_1}\left[M(t_1')_I\right] = 0 \Rightarrow \beta_1 = \rho\frac{C_Y}{C_X}$$

After replacing value of β_1 in (5.4) , we obtained

$$\left[M(t_1')_I\right]_{Min} = \left[\delta_1 - (\delta_2 - \delta_3)\rho^2\right] S_Y^2$$

(ii) Similar to (i), we proceed for (5.5), we have

$$\frac{d}{d\beta_1}\left[M(t_1')_{II}\right] = 0 \Rightarrow \beta_1 = \delta_5(\delta_3 + \delta_5)^{-1}\rho\frac{C_Y}{C_X}$$

Hence, $$\left[M(t_1')_{II}\right]_{Min} = \left[\delta_4 - (\delta_3 + \delta_5)^{-1}\delta_5^2\rho^2\right] S_Y^2$$

Theorem 5.2

(5) The estimator t_2' in terms of e_1, e_2, e_3 and e_3' is

$$t_2' = \bar{Y}\left[1 + e_1 + \beta_2\left\{e_3 - e_2 + e_1 e_3 - e_1 e_2 - \beta_2 e_2 e_3 + \frac{\beta_2 + 1}{2}e_2^2 + \frac{\beta_2 - 1}{2}e_3^2\right\}\right] \quad (5.9)$$

(6) The bias of t_2' under design *I* and *II* respectively is

(i) $$B(t_2')_I = \bar{Y}\beta_2(\delta_1 - \delta_2)\left(\frac{\beta_2 + 1}{2}C_X^2 - \rho C_Y C_X\right) \quad (5.10)$$

(ii) $$B(t_2')_{II} = \bar{Y}\beta_2(\delta_4 - \delta_5)\left[\frac{1}{2}(\beta_2 + 1)C_X^2 - \rho C_Y C_X\right] \quad (5.11)$$

(7) Mean squared error of $t_2^{'}$ under design *I* and *II* respectively is:

(i) $$M\left(t_2^{'}\right)_I = \overline{Y}^2\left[\delta_1 C_Y^2 + \left(\delta_1 - \delta_2\right)\left(\beta_2^2 C_X^2 - 2\beta_2 \rho C_Y C_X\right)\right] \quad (5.12)$$

(ii) $$M\left(t_2^{'}\right)_{II} = \overline{Y}^2\left[\delta_4 C_Y^2 + \left(\delta_4 - \delta_5\right)\left\{\beta_2^2 C_X^2 - 2\beta_2 \rho C_Y C_X\right\}\right] \quad (5.13)$$

(8) The minimum m.s.e. of $t_2^{'}$ is

(i) $$\left[M\left(t_2^{'}\right)_I\right]_{Min} = \left[\delta_1 - \left(\delta_1 - \delta_2\right)\rho^2\right] S_Y^2 \quad \text{when } \beta_2 = \rho\frac{C_Y}{C_X} \quad (5.14)$$

(ii) $$\left[M\left(t_2^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - \left(\delta_4 - \delta_5\right)\rho^2\right] S_Y^2 \quad \text{when } \beta_2 = \rho\frac{C_Y}{C_X} \quad (5.15)$$

Theorem 5.3

(9) The estimator $t_3^{'}$ in terms of e_1, e_2, e_3 and $e_3^{'}$ is

$$t_3^{'} = \overline{Y}\left[1 + e_1 + \beta_3\left\{e_3^{'} - e_2 - e_1 e_2 + e_1 e_3^{'} - \beta_3 e_2 e_3^{'} + \frac{\beta_3 + 1}{2} e_2^2 + \frac{\beta_2 - 1}{2} e_3^{'2}\right\}\right] \quad (5.16)$$

(10) Bias of $t_3^{'}$ under design *I* and *II* respectively is:

(i) $$B\left(t_3^{'}\right)_I = \overline{Y}\beta_3\left(\delta_1 - \delta_3\right)\left(\frac{\beta_3 + 1}{2} C_X^2 - \rho C_Y C_X\right) \quad (5.17)$$

(ii) $$B\left(t_3^{'}\right)_{II} = \overline{Y}\beta_3\left[\frac{1}{2}\left\{\beta_3\left(\delta_4 + \delta_3\right) - \left(\delta_3 - \delta_4\right)\right\} C_X^2 - \delta_4 \rho C_Y C_X\right] \quad (5.18)$$

(11) Mean squared error of $t_3^{'}$ is:

(i) $$M\left(t_3^{'}\right)_I = \overline{Y}^2\left[\delta_1 C_Y^2 + \left(\delta_1 - \delta_3\right)\left(\beta_3^2 C_X^2 - 2\beta_3 \rho C_Y C_X\right)\right] \quad (5.19)$$

(ii) $$M\left(t_3^{'}\right)_{II} = \overline{Y}^2\left[\delta_4 C_Y^2 + \left(\delta_3 + \delta_4\right)\beta_3^2 C_X^2 - 2\delta_4 \beta_3 \rho C_Y C_X\right] \quad (5.20)$$

(12) The minimum m.s.e. of $t_3^{'}$ is:

(i) $$\left[M\left(t_3^{'}\right)_I\right]_{\min} = \left[\delta_1 - \left(\delta_1 - \delta_3\right)\rho^2\right] S_y^2 \quad \text{when } \beta_3 = \rho\frac{C_Y}{C_X} \quad (5.21)$$

(ii) $$\left[M\left(t_3^{'}\right)_{II}\right]_{\min} = \left[\delta_4 - \delta_4^2\left(\delta_3 + \delta_4\right)^{-1}\rho^2\right] S_y^2 \quad \text{when } \beta_3 = \delta_4\left(\delta_3 + \delta_4\right)^{-1}\rho\frac{C_Y}{C_X} \quad (5.22)$$

6. Comparisons

In this section we derived the conditions under which the suggested estimators are superior to the Ahmed et al. (2006) over design *I* and *II*.

(1) $$\Delta_1 = \min\left[M\left(t_1\right)\right] - \min\left[M\left(t_1^{'}\right)_I\right] = \left[\frac{1}{n^{'}} - \frac{1}{N}\right] S_Y^2 - 2\left[\frac{1}{n^{'}} - \frac{1}{N}\right]\rho^2 S_y^2$$

$(t_1')_I$ is better than t_1, if $\Delta_1 > 0 \quad \Rightarrow -\frac{1}{2} < \rho < \frac{1}{2}$

(2) $\Delta_2 = \min[M(t_1)] - \min[M(t_1')_{II}] = [\delta_6 - \delta_4]\, S_Y^2 - [\delta_7 - (\delta_3 + \delta_5)^{-1}\delta_5^2]\, \rho^2 S_Y^2$

$(t_1')_{II}$ is better than t_1, if $\Delta_2 > 0$

$$\Rightarrow \rho^2 < \frac{(\delta_6 - \delta_4)(\delta_3 + \delta_5)}{[\delta_7(\delta_3 + \delta_5) - \delta_5^2]} \quad \Rightarrow -A < \rho < A$$

(3) $\Delta_3 = \min[M(t_2)] - \min[M(t_2')_I] = \left(\frac{1}{n'} - \frac{1}{N}\right) S_Y^2$

$(t_2')_I$ is better than t_2 if $\Delta_3 > 0$

$$\Rightarrow \left(\frac{1}{n'} - \frac{1}{N}\right) > 0 \Rightarrow N - n' > 0 \quad \Rightarrow n' < N$$

which is always true.

(4) $\Delta_4 = \min[M(t_2)] - \min[M(t_2')_{Ii}] = \left[\frac{1}{N - n'} - \frac{1}{N}\right] S_Y^2$

$(t_2')_{II}$ is better than t_2 if $\Delta_4 > 0$

\ $$\Rightarrow \left[\frac{1}{N - n'} - \frac{1}{N}\right] S_Y^2 > 0 \quad \Rightarrow n' > 0$$

which is always true.

(5) $\Delta_5 = \min[M(t_3)] - \min[M(t_3')_I] = \left[\frac{1}{n'} - \frac{1}{N}\right] S_Y^2 - \left(\frac{2}{n'}\right) \rho^2 S_y^2$

$(t_3')_I$ is better than t_3 if $\Delta_5 > 0 \Rightarrow \rho^2 < \frac{1}{2}\left[\frac{N - n'}{N}\right]$

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

(6) $\Delta_6 = \min[M(t_3)] - \min[M(t_3')_{II}] = [\delta_8 - \delta_4]\, S_Y^2 - [\delta_8 - (\delta_3 + \delta_4)^{-1}\delta_4^2]\, \rho^2 S_Y^2$

$(t_3')_{II}$ is better than t_3, if $\Delta_6 > 0$

$$\Rightarrow \rho^2 < \frac{(\delta_8 - \delta_4)(\delta_3 + \delta_4)}{[\delta_8(\delta_3 + \delta_4) - \delta_4^2]} \quad \Rightarrow -B < \rho < B$$

7. Numerical Illustrations

We consider two populations A and B, first one is the artificial population of size N = 200 [source Shukla et al. (2009)] and another one is from Ahmed et al. (2006) with the following parameters:

Table 7.0: Parameters of Populations A and B

Population	N	$\overline{Y}$	$\overline{X}$	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 5) and Ahmed et al. (2006) methods (see Appendix A) are given in table 7.1, 7.2 and 7.3 for population A and B respectively.

Table 7.1: Bias and MSE for Population A

Estimators	**Design I**		**Design II**	
	Bias	**MSE**	**Bias**	**MSE**
$t_1^{'}$	-0.00180934	36.990998	0.123403	36.78069
$t_2^{'}$	0.094991	10.4174764	0.94991	12.31328
$t_3^{'}$	0.09318118	10.91417774	1.843024	11.29167

Table 7.2: Bias and MSE for Population B

Estimators II	**Design I**		**Design II**	
	Bias	**MSE**	**Bias**	**MSE**
$t_1^{'}$	0.25646	22261.45	0.378708	22339.4
$t_2^{'}$	14.80248	16403.58	14.80248	16518.98
$t_3^{'}$	15.05895	16300.3	8.94385	16384.03

Table 7.3: Bias and MSE for Population A and B for Ahmed et al. (2006)

Estimators	**Population A**		**Population B**	
	Bias	**MSE**	**Bias**	**MSE**
t_1	0.010856	35.83645	15.23273	22319.77
t_2	0.094991	12.73984	14.80248	16531.89
t_3	0.105847	9.759633	15.23273	16358.62

The sampling efficiency of suggested estimators under design I and II over Ahmed et al. (2006) is defined as:

$$E_i = \frac{Opt\left[M\left(t_i^{'}\right)_j\right]}{Opt\left[M\left(t_i\right)\right]}; \qquad i = 1,2,3; \qquad j = I, II \qquad \ldots(7.1)$$

The efficiency for population A and B, respectively given in table 7.4.

Table 7.4: Efficiency for Population A and B over Ahmed et al. (2006)

Efficiency	Population A		Population B	
	Design I	Design II	Design I	Design II
E_1	1.032217	1.026349	0.997367	1.000879
E_2	0.817709	0.966518	0.992239	0.999219
E_3	1.118298	1.156977	0.996435	1.001553

8. Discussion

The idea of two-phase sampling is used while considered that the auxiliary population mean is unknown. Some strategies are suggested in Section 4 and the estimator of population mean derived. Properties of derived estimators like bias and m.s.e are discussed in the Section 5. The optimum value of parameters of suggested estimators is obtained as well in same section. Ahmed et al. (2006) estimators are considered for comparison purpose and two populations A and B considered for numerical study first one from Shukla et al. (2009) and another one is Ahmed et al. (2006). The sampling efficiency of suggested estimator under design *I* and *II* over Ahmed et al. (2006) is obtained and suggested strategy is found very close with Ahmed et al. (2006) when $\bar{X}$ is not known.

9. Conclusion

The proposed estimators are useful when some observations are missing in the sampling and population mean of auxiliary information is unknown. Obviously from Table 7.1 and 7.2, all suggested estimators are better in design *I* than design *II* i.e. the design *I* is better than design *II*. Table 7.3 shows bias and m.s.e for population A and B for Ahmed et al. (2006). From table 7.4 it is obvious that the suggested strategies are very close with Ahmed et al. (2006).

Acknowledgement

Authors are thankful to the Editorial Board of JRSS and referees for recommending the manuscript for publication.

References

1. Ahmed, M.S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006): Estimation of a population mean using different imputation methods, Statistics in Transition, 7(6), p. 1247-1264.
2. Kalton, G., Kasprzyk, D. and Santos, R. (1981): Issues of non-response and imputation in the Survey of Income and Program Participation. Current Topics in Survey Sampling, (D. Krewski, R. Platek and J.N.K. Rao, eds.), p. 455-480, Academic Press, New York.

3. Lee, H., Rancourt, E. and Sarndal,C. E. (1994):Experiments with variance estimation from survey data with imputed values. Journal of official Statistics, 10(3). p. 231-243.
4. Lee, H., Rancourt, E. and Sarndal,C. E. (1995): Variance estimation in the presence of imputed data for the generalized estimation system. Proc. of the American Statist. Assoc. (Social Survey Research Methods Section), p. 384-389.
5. Rao, J. N. K. and Sitter, R. R. (1995): Variance estimation under two-phase sampling with application to imputation for missing data, Biometrica, 82, p. 453-460.
6. Sande, I. G. (1979): A personal view of hot deck approach to automatic edit and imputation. Journal Imputation Procedures. Survey Methodology, 5, p. 238-246.
7. Shukla, D. (2002): F-T estimator under two-phase sampling, Metron, 59, 1-2, p. 253-263.
8. Shukla, D. and Thakur, N. S. (2008): Estimation of mean with imputation of missing data using factor-type estimator, Statistics in Transition, 9(1), p.33-48.
9. Shukla, D., Thakur, N. S., Pathak, S. and Rajput D. S. (2009): Estimation of mean with imputation of missing data using factor- type estimator in two-phase sampling, Statistics in Transition, 10(3), p. 397-414.
10. Shukla, D., Thakur, N. S., Thakur, D. S. (2009):Utilization of non-response auxiliary population mean in imputation for missing observations, Journal of Reliability and Statistical Studies, 2(1), p. 28-40.
11. Shukla, D., Thakur, N. S., Thakur, D. S. and Pathak, S. (2011): Linear combination based imputation method for missing data in sample, International Journal of Modern Engineering Research (IJMER), 1(2), p. 580-596.
12. Singh, S. and Horn, S. (2000): Compromised imputation in survey sampling, Metrika, 51, p. 266-276.

Appendix – A

Proposed Methods of Ahmed et al. (2006)

Ahmed et al. (2006) proposed some imputation methods and derived their properties. Authors are discussing with three methods of them. Let y_{ji} denotes the i^{th} available observation for the j^{th} imputation and β_i , $i = 1,2,3$ is a suitably chosen constant, such that the variance the resultant estimator is minimum. Imputation methods are :

(1)
$$y_{1i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{\bar{x}}}\right)^{\beta_1} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{1}$$

Under this method, the point estimator of $\bar{Y}$ is $t_1 = \bar{y}_r\left(\dfrac{\bar{X}}{\bar{\bar{x}}}\right)^{\beta_1}$ (2)

Theorem: The bias, mean squared error and minimum mean squared error at $\beta_1 = \rho\dfrac{C_Y}{C_X}$ of t_1 is given by

(i)
$$B(t_1)_I = \bar{Y}\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{\beta_1(\beta_1+1)}{2}C_X^2 - \beta_1\rho C_Y C_X\right) \tag{3}$$

(ii)
$$M(t_1)_I \approx \bar{Y}^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)C_Y^2 + \beta_1^2\left(\frac{1}{n}-\frac{1}{N}\right)C_X^2 - 2\beta_1\left(\frac{1}{n}-\frac{1}{N}\right)\rho C_Y C_X\right] \tag{4}$$

(iii)
$$M(t_1)_{\min} \approx \left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2 - \left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{XY}^2}{S_X^2} \tag{5}$$

(2)
$$y_{2i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{x}}{\bar{\bar{x}}_r}\right)^{\beta_2} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{6}$$

Under this method, the point estimator of $\bar{Y}$ is $t_2 = \bar{y}_r\left(\dfrac{\bar{x}}{\bar{\bar{x}}_r}\right)^{\beta_2}$ (7)

Theorem: The bias, mean squared error and minimum mean squared error at $\beta_2 = \rho\dfrac{C_Y}{C_X}$ of t_2 is given by

(i)
$$B(t_2) = \left(\frac{1}{r}-\frac{1}{n}\right)\bar{Y}\left(\frac{\beta_2(\beta_2+1)}{2}C_X^2 - \beta_2\rho C_Y C_X\right) \tag{8}$$

(ii)
$$M(t_2)_I \approx \bar{Y}^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)C_Y^2 + \beta_2^2\left(\frac{1}{r}-\frac{1}{n}\right)C_X^2 - 2\beta_2\left(\frac{1}{r}-\frac{1}{n}\right)\rho C_Y C_X\right] \tag{9}$$

(iii) $$M(t_2)_{\min} \approx \left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2 - \left(\frac{1}{r}-\frac{1}{n}\right)\frac{S_{XY}^2}{S_X^2} \qquad (10)$$

(3) $$y_{3i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_r}\right)^{\beta_3} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \qquad (11)$$

Under this method, the point estimator of $\bar{Y}$ is $$t_3 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\beta_3} \qquad (12)$$

Theorem: The bias, mean squared error and minimum mean squared error at $\beta_3 = \rho\frac{C_Y}{C_X}$ of t_3 is given by

(i) $$B(t_3) = \left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}\left(\frac{\beta_3(\beta_3+1)}{2}C_X^2 - \beta_3\rho C_Y C_X\right) \qquad (13)$$

(ii) $$M(t_3) \approx \left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}^2\left[C_Y^2 + \beta_3^2 C_X^2 - 2\beta_3\rho C_Y C_X\right] \qquad (14)$$

(iii) $$M(t_3)_{\min} \approx \left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2\left(1-\rho^2\right) \qquad .(15)$$