SOME IMPUTATION METHODS IN DOUBLE SAMPLING SCHEME FOR ESTIMATION OF POPULATION MEAN

Narendra Singh Thakur, Kalpana Yadav
Centre for Mathematical Sciences (CMS), Banasthali University, Rajasthan-304022 INDIA
and
Sharad Pathak
Department of Mathematics and Statistics, Dr. H. S. Gour Central Univesity, Sagar (M.P.) INDIA

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's methods of imputation, F-T methods 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's estimators.

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

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) discussed 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 $U = (U_1, U_2, U_3,\ldots,U_N)$ be the finite population of size N and the character under study be denoted by 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 ($n < n'$) is 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 by 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; $\delta_1 = \left(\frac{1}{r} - \frac{1}{n'}\right)$; $\delta_2 = \left(\frac{1}{n} - \frac{1}{n'}\right)$; $\delta_3 = \left(\frac{1}{n'} - \frac{1}{N}\right)$; $\delta_4 = \left(\frac{1}{r} - \frac{1}{N-n'}\right)$; $\delta_5 = \left(\frac{1}{n} - \frac{1}{N-n'}\right)$; $f_1 = \frac{r}{n}$,

$$C = \frac{(\delta_9 - \delta_4)(\delta_3 + \delta_5)}{\left[\delta_{10}(\delta_3 + \delta_5) - \delta_5^2\right]}; \; D = \frac{(\delta_{11} - \delta_4)(\delta_3 + \delta_4)}{\left[\delta_{11}(\delta_3 + \delta_4) - \delta_4^2\right]}.$$

3. LARGE SAMPLE APPROXIMATIONS:

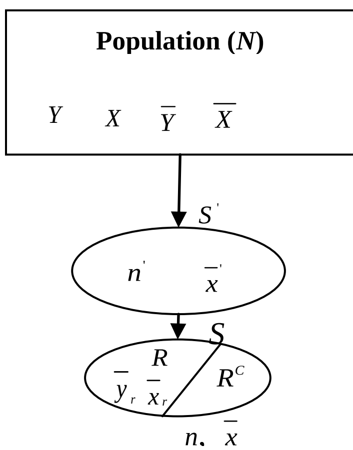

Fig. 1 [Design I]

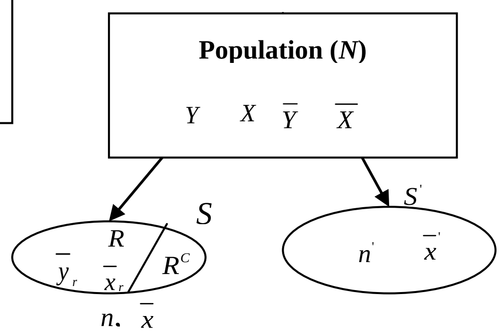

Fig. 2 [Design II]

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	$E(e_i)$	$E(e_3')$	$E(e_1^2)$	$E(e_2^2)$	$E(e_3^2)$	$E(e'^2_3)$
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 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:

(1) $$y'_{4i} = \begin{cases} y_i & if \quad i \in R \\ \left[\bar{y}_r + b_1\left(x_i - \bar{x}_r\right)\right] & if \quad i \in R^C \end{cases} \qquad \ldots(4.1)$$

under this strategy, the point estimator of $\bar{Y}$ is given by $t'_4 = \bar{y}_r + b_1\left(\bar{x} - \bar{x}_r\right)$ …(4.2)

(2) $$y'_{5i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r + \dfrac{b_2}{(1-f_1)}\left(\bar{x}' - \bar{x}\right) & if \quad i \in R^C \end{cases} \qquad \ldots(4.3)$$

under this , the estimator of $\bar{Y}$ is $t'_5 = \bar{y}_r + b_2\left(\bar{x}' - \bar{x}\right)$ …(4.4)

(3) $$y'_{6i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r + \dfrac{b_3}{(1-f_1)}\left(\bar{x}' - \bar{x}_r\right) & if \quad i \in R^C \end{cases} \qquad \ldots(4.5)$$

hence the estimator of $\bar{Y}$ is $t'_6 = \bar{y}_r + b_3\left(\bar{x}' - \bar{x}_r\right)$ …(4.6)

5. BIAS AND MEAN SQUARED ERROR 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 $t = I, II$, then the bias and m.s.e of t'_4, t'_5 and t'_6 are derived in the following theorems. 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'_4 in terms of e_i; $i = 1,2,3$ and e'_3 could be expressed:

$$t'_4 = \bar{Y}(1+e_1) + b_1\bar{X}(e_3 - e_2) \qquad \ldots(5.1)$$

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: $t'_4 = \bar{y}_r + b_1\left(\bar{x} - \bar{x}_r\right) \quad = \bar{Y}(1+e_1) + b_1\bar{X}\left(e_3 - e_2\right)$

(2) Bias of t'_4 under design *I* and *II* is:

(i) $B\left[t'_4\right]_I = 0$ …(5.2)

(ii) $B\left[t'_4\right]_{II} = 0$ …(5.3)

Proof:

(i) $B\left(t'_4\right)_I = E\left[t'_4 - \bar{Y}\right]_I = \bar{Y} - \bar{Y} = 0$

(ii) $B\left(t'_4\right)_{II} = E\left[t'_4 - \bar{Y}\right]_{II} = \bar{Y} - \bar{Y} = 0$

(3) The variance of t'_4, under design *I* and *II*, upto first order of approximation could be written as:

(i) $V\left(t'_4\right)_I = \delta_1 S_Y^2 + (\delta_1 - \delta_2)\left(b_1^2 S_X^2 - 2b_1\rho S_Y S_X\right)$ …(5.4)

(ii) $V\left(t'_4\right)_I = \delta_4 S_Y^2 + (\delta_4 - \delta_5)\left(b_1^2 S_X^2 - 2b_1\rho S_Y S_X\right)$ …(5.5)

Proof: $V\left(t'_4\right) = E\left[t'_4 - \bar{Y}\right]^2 = E\left[\bar{Y}e_1 + b_1\bar{X}\left(e_3 - e_2\right)\right]^2$

$$= E\left[\bar{Y}^2 e_1^2 + b_1^2\bar{X}^2\left(e_3 - e_2\right)^2 + 2b_1\bar{Y}\bar{X}\left(e_3 - e_2\right)e_1\right]$$

$$= E\left[\bar{Y}^2 e_1^2 + b_1^2 \bar{X}^2 \left(e_3^2 + e_2^2 - 2e_2 e_3\right) + 2b_1 \bar{Y}\bar{X}\left(e_1 e_3 - e_1 e_2\right)\right] \quad \ldots(5.6)$$

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

$$V\left(t_4^{'}\right)_I = \left[\ \bar{Y}^2 \delta_1 C_Y^2 + b_1^2 \bar{X}^2 \left(\delta_2 C_X^2 + \delta_1 C_X^2 - 2\delta_2 C_X^2\right) + 2b_1 \bar{Y}\bar{X}\left(\delta_2 \rho C_Y C_X - \delta_1 \rho C_Y C_X\right)\ \right]$$

$$= \left[\delta_1 S_Y^2 + b_1^2 S_X^2 \left(\delta_1 - \delta_2\right) - 2b_1\left(\delta_1 - \delta_2\right)\rho S_Y S_X\ \right]$$

$$= \left[\delta_1 S_Y^2 + \left(\delta_1 - \delta_2\right)\left\{b_1^2 S_X^2 - 2b_1 \rho S_Y S_X\right\}\right]$$

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

$$V\left(t_4^{'}\right)_{II} = \left[\bar{Y}^2 \delta_1 C_Y^2 + b_1^2 \bar{X}^2 \left(\delta_5 C_X^2 + \delta_4 C_X^2 - 2\delta_5 C_X^2\right) + 2b_1 \bar{Y}\bar{X}\left(\delta_5 \rho C_Y C_X - \delta_4 \rho C_Y C_X\right)\right]$$

$$= \left[\delta_4 S_Y^2 + b_1^2 S_X^2 \left(\delta_4 - \delta_5\right) - 2b_1\left(\delta_4 - \delta_5\right)\rho S_Y S_X\ \right]$$

(4) The minimum variance of the $t_4^{'}$ is

(i) $$\left[V\left(t_4^{'}\right)_I\right]_{Min} = \left[\delta_1 - \left(\delta_1 - \delta_2\right)\rho^2\right] S_Y^2 \quad \text{when} \quad b_1 = \rho \frac{S_Y}{S_X} \quad \ldots(5.7)$$

(ii) $$\left[V\left(t_4^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - \left(\delta_4 - \delta_5\right)\rho^2\right] S_Y^2 \quad \text{when} \quad b_1 = \rho \frac{S_Y}{S_X} \quad \ldots(5.8)$$

Proof:

(i) By differentiating (5.4) with respect to b_1 and equate to zero, we get

$$\frac{d}{db_1}\left[V\left(t_4^{'}\right)_I\right] = 0 \ \Rightarrow b_1 = \rho \frac{S_Y}{S_X}$$

After replacing the value of b_1 in (5.4), we obtained

$$\left[V\left(t_4^{'}\right)_I\right]_{Min} = \delta_1 S_Y^2 + \left(\delta_1 - \delta_3\right)\left(\rho^2 S_Y^2 - 2\rho^2 S_Y^2\right)$$

$$= \left[\delta_1 - \left(\delta_1 - \delta_2\right)\rho^2\right] S_Y^2$$

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

$$\frac{d}{db_1}\left[V\left(t_4^{'}\right)_{II}\right] = 0 \ \Rightarrow b_1 = \rho \frac{S_Y}{S_X}$$

After replacing the value of b_1 in (5.5), we obtained

$$\left[V\left(t_4^{'}\right)_I\right]_{Min} = \delta_4 S_Y^2 + \left(\delta_4 - \delta_5\right)\left\{\rho^2 S_Y^2 - 2\rho^2 S_Y^2\right\}$$

$$= \left[\delta_4 - \left(\delta_4 - \delta_5\right)\rho^2\right] S_Y^2$$

Theorem 5.2:

(5) The estimator $t_5^{'}$ in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$t_5^{'} = \bar{Y}\left(1 + e_1\right) + b_2 \bar{X}\left(e_3^{'} - e_3\right) \quad \ldots(5.9)$$

(6) The bias estimator $t_5^{'}$, under design *I* and *II* respectively is

(i) $$B\left[t_5^{'}\right]_I = 0 \quad \ldots(5.10)$$

(ii) $$B\left[t_5^{'}\right]_{II} = 0 \quad \ldots(5.11)$$

(7) The variance of $t_5^{'}$, under design *I* and *II* respectively is:

(i) $$V\left(t_5^{'}\right)_I = \delta_1 S_Y^2 + \left(\delta_2 - \delta_3\right)\left(b_2^2 S_X^2 - 2b_2 \rho S_Y S_X\right) \quad \ldots(5.12)$$

(ii) $$V\left(t_5^{'}\right)_{II} = \delta_4 S_Y^2 + \left(\delta_3 + \delta_5\right)b_2^2 S_X^2 - 2b_2 \delta_5 \rho S_Y S_X \quad \ldots(5.13)$$

(8) The minimum variance of the $t_5^{'}$ is

(i) $\left[V\left(t_5^{'}\right)_I\right]_{Min} = \left[\delta_1 - (\delta_2 - \delta_3)\rho^2\right] S_Y^2$ when $b_2 = \rho \dfrac{S_Y}{S_X}$ …(5.14)

(ii) $\left[V\left(t_5^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - \delta_5^2(\delta_3 + \delta_5)^{-1}\rho^2\right] S_Y^2$ when $b_2 = \left(\dfrac{\delta_5}{\delta_3 + \delta_5}\right)\rho \dfrac{S_Y}{S_X}$ …(5.15)

Theorem 5.3:

(9) The estimator $t_6^{'}$ in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$t_6^{'} = \overline{Y}(1 + e_1) + b_3 \overline{X}\left(e_3^{'} - e_2\right) \quad \text{…(5.16)}$$

(10) The bias estimator $t_6^{'}$, under design *I* and *II* respectively is:

(i) $B\left[t_6^{'}\right]_I = 0$ …(5.17)

(ii) $B\left[t_6^{'}\right]_{II} = 0$ …(5.43)

(11) The variance of $t_6^{'}$, under F_1 and F_2 is

(i) $V\left(t_6^{'}\right)_I = \delta_1 S_Y^2 + (\delta_2 - \delta_3)\left(b_3^2 S_X^2 - 2b_3\rho S_Y S_X\right)$ …(5.18)

(ii) $V\left(t_6^{'}\right)_{II} = \delta_4 S_Y^2 + (\delta_3 + \delta_4)b_3^2 S_X^2 - 2b_3\delta_4\rho S_Y S_X$ …(5.19)

(12) The minimum variance of the $t_6^{'}$ is

(i) $\left[V\left(t_6^{'}\right)_I\right]_{Min} = \left[\delta_1 - (\delta_1 - \delta_3)\rho^2\right] S_Y^2$ when $b_3 = \rho \dfrac{S_Y}{S_X}$ …(5.20)

(ii) $\left[V\left(t_6^{'}\right)_{II}\right]_{Min} = \left[\delta_4 - \delta_4^2(\delta_3 + \delta_4)^{-1}\rho^2\right] S_Y^2$ when $b_3 = \left(\dfrac{\delta_4}{\delta_3 + \delta_4}\right)\rho \dfrac{S_Y}{S_X}$ …(5.21)

6. COMPARISIONS:

(1) $\Delta_1 = \min\left[V(t_4)\right] - \min\left[V\left(t_4^{'}\right)_I\right] = \left[\dfrac{1}{n^{'}} - \dfrac{1}{N}\right] S_Y^2$

$\left(t_4^{'}\right)_I$ is better than t_4, if $\Delta_1 > 0 \Rightarrow \left[\dfrac{N - n^{'}}{n^{'}N}\right] > 0 \Rightarrow N - n^{'} > 0 \Rightarrow n^{'} < N$

which is always true.

(2) $\Delta_2 = \min\left[V(t_4)\right] - \min\left[V\left(t_4^{'}\right)_{II}\right] = \left[\dfrac{1}{N - n^{'}} - \dfrac{1}{N}\right] S_Y^2$

$\left(t_4^{'}\right)_{II}$ is better than t_4, if $\Delta_2 > 0 \Rightarrow \left[\dfrac{N - N + n^{'}}{N\left(N - n^{'}\right)}\right] > 0 \Rightarrow n^{'} > 0$

which is always true.

(3) $\Delta_3 = \min\left[V(t_5)\right] - \min\left[V\left(t_5^{'}\right)_I\right] = \left[\dfrac{1}{n^{'}} - \dfrac{1}{N}\right] S_Y^2 + \left[\dfrac{2}{N} - \dfrac{2}{n^{'}}\right]\rho^2 S_y^2$

$\left(t_5^{'}\right)_I$ is better than t_5, if $\Delta_3 > 0 \Rightarrow -\dfrac{1}{2} < \rho < \dfrac{1}{2}$

(4) $\Delta_4 = \min\left[V(t_5)\right] - \min\left[V\left(t_5^{'}\right)_{II}\right] = \left[\delta_9 - \delta_4\right] S_Y^2 - \left[\delta_{10} - (\delta_3 + \delta_5)^{-1}\delta_5^2\right]\rho^2 S_Y^2$

$\left(t_5^{'}\right)_{II}$ is better than t_5, if $\Delta_4 > 0 \Rightarrow \rho^2 < \dfrac{(\delta_9 - \delta_4)(\delta_3 + \delta_5)}{\left[\delta_{10}(\delta_3 + \delta_5) - \delta_5^2\right]} \Rightarrow -C < \rho < C$

(5) $\Delta_5 = \min\left[V(t_6)\right] - \min\left[V\left(t_6^{'}\right)_I\right] = \left[\dfrac{1}{n^{'}} - \dfrac{1}{N}\right] S_Y^2 + \left[\dfrac{2}{N} - \dfrac{2}{n^{'}}\right]\rho^2 S_y^2$

$(t_6^{'})_{I}$ is better than t_6, if $\Delta_5 > 0$ $\Rightarrow 2\left[\frac{1}{n^{'}}-\frac{1}{N}\right]\rho^2 < \left(\frac{1}{n^{'}}-\frac{1}{N}\right)$ $\Rightarrow -\frac{1}{2} < \rho < \frac{1}{2}$

(6) $\Delta_6 = \min\left[V(t_6)\right] - \min\left[V(t_6^{'})_{II}\right] = [\delta_{11}-\delta_4]\, S_Y^2 - \left[\delta_{11} - (\delta_3+\delta_4)^{-1}\delta_4^2\right] \rho^2 S_Y^2$

$(t_6^{'})_{II}$ is better than t_6, if $\Delta_6 > 0$ $\Rightarrow \rho^2 < \frac{(\delta_{11}-\delta_4)(\delta_3+\delta_4)}{\left[\delta_{11}(\delta_3+\delta_4)-\delta_4^2\right]}$ $\Rightarrow -D < \rho < D$

7. NUMERICAL ILLUSTRATIONS:

We considered 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	$\bar{Y}$	$\bar{X}$	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 for suggested estimators under design *I* and *II* , using the expressions of bias and M.S.E. derived in Section 5 for suggested estimators are shown in table 7.1 and 7.2 respectively. The bias and M.S.E. for Ahmed's estimators (see Appendix A) are displayed in table 7.3 for population A and B respectively.

Table 7.1 Bias and MSE (Population A)

Estimators	DESIGN *I*		DESIGN *II*	
	Bias	MSE	Bias	MSE
$t_4^{'}$	0	10.41747	0	12.31328
$t_5^{'}$	0	36.99100	0	36.78069
$t_6^{'}$	0	10.91418	0	11.29167

Table 7.2 Bias and MSE (Population B)

Estimators	DESIGN *I*		DESIGN *II*	
	Bias	MSE	Bias	MSE
$t_4^{'}$	0	16403.58	0	16518.98
$t_5^{'}$	0	22261.45	0	22339.40
$t_6^{'}$	0	16300.30	0	16384.03

Table 7.3 Bias and MSE for Ahmed's Estimators (Population A and B)

Estimators	Population A		Population B	
	Bias	MSE	Bias	MSE
t_4	0	12.73984	0	16531.89
t_5	0	35.83645	0	22319.77
t_6	0	9.759633	0	16358.62

The sampling efficiency of suggested estimators under design *I* and *II* over Ahmed's estimators is defined as:

$$E_i = \frac{Opt\left[M\left(t_i^{'}\right)_j\right]}{Opt\left[M\left(t_i\right)\right]}; \qquad i = 4,5,6; \qquad j = I, II \qquad \ldots(7.1)$$

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

Table 7.4 Efficiency of Suggested Estimators in Design *I* and *II* over Ahmed's Estimators

Efficiency	Population A		Population B	
	Design *I*	Design *II*	Design *I*	Design *II*
E_4	0.817709	0.966518	0.992239	0.999219
E_5	1.032217	1.026349	0.997387	1.000879
E_6	1.118298	1.156977	0.996435	1.001553

8. DISCUSSIONS:

The idea of two-phase sampling is used while considering the auxiliary population mean is unknown. Some strategies are suggested for missing observations in Section 4 and the estimators of population mean are derived. Properties of estimators like bias and m.s.e are discussed in the Section 5 and the optimum value of parameters for minimum mean squared error is obtained as well in the same section. Ahmed's estimators are considered for relative comparison. 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's estimators is obtained and suggested strategy is found very close with Ahmed et al. (2006) when $\overline{X}$ unknown.

9. CONCLUSIONS:

The proposed estimators are useful when some observations are missing in the sample 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*. The table 7.4 shows that the suggested estimators $t_5^{'}$ and $t_6^{'}$ are very close with Ahmed's estimators and may be used to estimate the population mean while population parameter of auxiliary information is unknown.

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, 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.), 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, 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 Ame. Stat. Asso. (Social Survey Res. Methods Sec.), 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, 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, 238-246.
7. **Shukla, D. (2002):** *F-T estimator under two-phase sampling,* Metron, 59, 1-2, 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, 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, 397-414.
10. **Shukla, D., Thakur, N. S., Thakur, D. S. (2009a):***Utilization of non-response auxiliary population mean in imputation for missing observations*, Journal of Reliability and Statistical Studies, 2, 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, 580-596.

12. **Singh, S. and Horn, S. (2000):** *Compromised imputation in survey sampling*, Metrika, 51, 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 $b_i \;, i = 1,2,3$ is a suitably chosen constant, such that the variance the resultant estimator is minimum. Imputation methods are :

(1) $$y_{4i} = \begin{cases} y_i & if \quad i \in R \\ \left[\bar{y}_r + b_1\left(\bar{x}_i - \bar{x}_r\right)\right] & if \quad i \in R^C \end{cases} \qquad \ldots(1)$$

under this strategy, the point estimator of $\bar{Y}$ is $t_4 = \bar{y}_r + b_1\left(\bar{x} - \bar{x}_r\right)$ …(2)

Theorem: The bias, variance and minimum variance at $b_1 = \rho\dfrac{S_{XY}}{S_X^2}$ of t_4 is given by

(i) $B[t_4] = 0$ …(3)

(ii) $V(t_4) = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)S_Y^2 + b_1^2\left(\dfrac{1}{r} - \dfrac{1}{n}\right)S_X^2 - 2b_1\left(\dfrac{1}{r} - \dfrac{1}{n}\right)S_{XY}$ …(4)

(iii) $V(t_4)_{\min} = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)S_Y^2 - \left(\dfrac{1}{r} - \dfrac{1}{n}\right)\dfrac{S_{XY}^2}{S_X^2}$ …(5)

(2) $$y_{5i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r + \dfrac{nb_2}{(n-r)}\left(\bar{X} - \bar{x}\right) & if \quad i \in R^C \end{cases} \qquad \ldots(6)$$

under this strategy, the point estimator of $\bar{Y}$ is $t_5 = \bar{y}_r + b_2\left(\bar{X} - \bar{x}\right)$ …(7)

Theorem: The bias, variance and minimum variance at $b_2 = \rho\dfrac{S_{XY}}{S_X^2}$ of t_5 is given by

(i) $B[t_5] = 0$ …(8)

(ii) $V(t_5) = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)S_Y^2 + b_2^2\left(\dfrac{1}{n} - \dfrac{1}{N}\right)S_X^2 - 2b_2\left(\dfrac{1}{n} - \dfrac{1}{N}\right)S_{XY}$ …(9)

(iii) $V(t_5)_{\min} = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)S_Y^2 - \left(\dfrac{1}{n} - \dfrac{1}{N}\right)\dfrac{S_{XY}^2}{S_X^2}$ …(10)

(3) $$y_{6i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r + \dfrac{nb_3}{(n-r)}\left(\bar{X} - \bar{x}_r\right) & if \quad i \in R^C \end{cases} \qquad \ldots(11)$$

under this , the estimator of $\bar{Y}$ is $t_6 = \bar{y}_r + b_3\left(\bar{X} - \bar{x}_r\right)$ …(12)

Theorem: The bias, variance and minimum variance at $b_3 = \rho\dfrac{S_{XY}}{S_X^2}$ of t_6 is given by

(i) $B[t_6] = 0$ …(13)

(ii) $V(t_6) = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)\left(S_Y^2 + b_3^2 S_X^2 - 2b_3 S_{XY}\right)$ …(14)

(iii) $V(t_6)_{\min} = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)S_Y^2\left(1 - \rho^2\right)$ …(15)