Some imputation methods for missing data in sample surveys

G. N. Singh* †, S. Maurya*, M. Khetan*and Cem Kadilar‡

Abstract

The present work suggests some imputation methods to deal with the problems of non-response in sample surveys. The imputation methods presented in this work lead to the precise estimation strategies of population mean. Empirical studies are carried out with the help of data borrowed from natural populations to show the superiorities of the suggested imputation methods over usual mean, ratio and regression methods of imputation in terms of the mean square error criterions. Suitable recommendations have been put forward for the survey practitioners.

Keywords: Imputation, non-response, auxiliary information, bias, mean square error.

2000 AMS Classification: 62D05

1. Introduction

The clinical or life savings drug testing experiments face the problems of missing data due to elimination of some of the experimental units during the course of experiments. Similarly in agricultural experiments, crops destroy due to some natural calamities or disease during the course of experiments. In demographic and socio-economic surveys, generally response from each unit in sample is not available due to various causes. Such incompleteness is known as non-response and if the appropriate information about the nature of non-response is not available, the conclusions concerning the population parameters may be spoiled.

In last couple of decades, significant advancements have been made to reduce the negative impact of non-response. Imputation is one which deals with the filling up method of incomplete data for adapting the standard analytic model in statistics. It is typically used when it is needed to substitute missing item values with certain fabricated values in a survey or census. To deal with the missing item values effectively [13], [14], [16] and [9] suggested imputation methods that make an incomplete data set structurally complete and its analysis simple. Imputation may also be carried out with the aid of an auxiliary variable if it is available. Some of the pioneer works which used information on an auxiliary variable under

*Department of Applied Mathematics, Indian School of Mines,Dhanbad 826004, India, Email: gnsingh_ism@yahoo.com

†Corresponding Author.

‡Department of Statistics, Hacettepe University, Beytepe Campus 06800, Ankara, Turkey, Email: kadilar@hacettepe.edu.tr

missing completely at random (MCAR) response mechanism were suggested by [10] ,[11] , [20],[22],[1],[4],[18],[21],[17],[19]and [2].

[15] advocated the use of multiple imputations to lessen the negative impact of missing data in more wise way. He showed multiple imputations provide a useful strategy for dealing with missing data by replacing each missing value with two or more acceptable fabricated values representing a distribution of possibilities. Motivated with this suggestion and in follow up we suggest some single and multiple imputations methods under MCAR response mechanism. The suggested imputation methods lead to some effective estimation procedures of population mean. Properties of the proposed imputation methods and subsequent estimation procedures have been examined and suitable recommendations are made.

2. Sample structure and notations

Consider $U = (U_1, U_2, U_3, ..., U_N)$ denote the finite population of size N and let y and x be the positively correlated study and auxiliary variables respectively. It is assumed that information on an auxiliary variable x is readily available for each unit of the population and we intend to estimate the population mean of the study variable y. Let a sample s of size n be drawn from the population under simple random sampling without replacement (SRSWOR) scheme and surveyed for study variable y but response from each sampled unit was not obtained which leads to the presence of non-response. Let r be the number of responding units out of sampled n units and the set of responding units is denoted by R and that of non-responding units by R^c . For sampled units $i \in R$, the values y_i are observed, while for the units $i \in R^c$,the y_i values are missing and respective imputed values are derived. We intend to develop some effective imputation methods with the aid of an auxiliary variable x, such that the value of x_i for unit U_i, is known and has positive value for each unit of the population. Hence onwards we use the following notations:
$\bar{Y}, \bar{X}$:The population means of the study and auxiliary variables y and x respectively.
S_y^2,S_x^2 :The population variances of the study and auxiliary variables y and x respectively.
C_y,C_x :The coefficients of variations of the study and auxiliary variables y and x respectively.
ρ_{yx}: The correlation coefficient between the study and auxiliary variables y and x.
$\bar{y}_r, \bar{x}_r$:The response means of the study and auxiliary variables y and x respectively.
$\bar{x}_n$:The sample mean of the auxiliary variable x based on the sample size n.

2.1. Proposed imputation methods and subsequent estimators. In this section, some more effective imputation methods and hence the corresponding estimators have been proposed under MCAR response mechanism. The derived resultant estimators have shown dominant performance over the existing methods of imputations and are more relevant for practical applications.

2.1.1. *Single imputation methods and subsequent estimators.* Following the MCAR response mechanism we suggest the following three single imputation methods for the missing values of the sample data.

(a) First method of imputation

The data after imputation takes the form,

$$y_{.i} = \begin{cases} y_i exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) & if \quad i \in R \\ \left(y_r + \hat{b}x_i - \hat{b}\bar{x}_i\right) exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) & if \quad i \in R^c \end{cases} \tag{2.1}$$

where $\hat{b} = \frac{s_{yx}(r)}{s_x^2(r)}$

Under the method of imputation discussed in equation (2.1), the point estimator of $\bar{Y}$ takes the following form

$$\tau_1 = \frac{1}{n}\sum_{i=1}^{n} y_{.i} = \frac{1}{n}\left[\sum_{i\in R} y_{.i} + \sum_{i\in R^c} y_{.i}\right] \tag{2.2}$$

which is simplified as

$$\tau_1 = \left[\bar{y}_r + \hat{b}\left(\bar{x}_n - \bar{x}_r\right)\right] exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) \tag{2.3}$$

(b) Second method of imputation

The data after imputation takes the form,

$$y_{.i} = \begin{cases} y_i exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) & if \quad i \in R \\ \left(\frac{\bar{y}_r}{\bar{x}_r}x_i\right) exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) & if \quad i \in R^c \end{cases} \tag{2.4}$$

Under the method of imputation described in equation (2.4), the point estimator of $\bar{Y}$ takes the following form

$$\tau_2 = \frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) \tag{2.5}$$

(c) Third method of imputation

The data after imputation takes the form,

$$y_{.i} = \begin{cases} y_i - \frac{n^2}{r^2}\bar{x}_n\hat{b} & if \quad i \in R \\ \left(\bar{y}_r + \frac{n}{n-r}\hat{b}\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) + \frac{n}{r}\hat{b}x_i\right) & if \quad i \in R^c \end{cases} \tag{2.6}$$

Under the method of imputation described in equation (2.6), the point estimator of $\bar{Y}$ takes the following form

$$\tau_3 = \bar{y}_r + \hat{b}\left[\left\{\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} - \bar{x}_r\right] \tag{2.7}$$

2.1.2. *Multiple imputations methods and resultant estimators.* In single imputation, the single value being imputed can reflect neither sampling variability about the actual value when one model for non-response is being considered nor additional uncertainty when more than one model is being entertained. Since, multiple imputations retain the virtues of single imputation and corrects its major flaws, therefore, we intend to use multiple imputations for each missing value in the sample of size n. The previously discussed methods of imputations have been considered to derive the imputed values for each missing value. After the generations of imputed values, complete data sets are produced and subsequently estimators based on sample of size n are reproduced. The final estimator of population mean $\bar{Y}$ is the average of estimates produced by imputation methods. Hence the final estimators of population mean $\bar{Y}$ based on the procedure of multiple imputations are considered as

$$\bar{y}_{MI_1} = \frac{1}{3}\left[\tau_1 + \tau_2 + \tau_3\right]$$

$$(2.8)\qquad \bar{y}_{MI_1} = \frac{1}{3}\left[\begin{array}{l} \left\{\bar{y}_r + \hat{b}\left(\bar{x}_n - \bar{x}_r\right) exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} \\ +\left\{\frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} \\ +\left\{\bar{y}_r + \hat{b}\left\{\left\{\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} - \bar{x}_r\right\}\right. \end{array}\right]$$

$$\bar{y}_{MI_2} = \frac{1}{2}\left[\tau_1 + \tau_2\right]$$

$$(2.9)\qquad \bar{y}_{MI_2} = \frac{1}{2}\left[\begin{array}{l} \left\{\bar{y}_r + \hat{b}\left(\bar{x}_n - \bar{x}_r\right) exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} \\ +\left\{\frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} \end{array}\right]$$

$$\bar{y}_{MI_3} = \frac{1}{2}\left[\tau_2 + \tau_3\right]$$

$$(2.10)\qquad \bar{y}_{MI_3} = \frac{1}{2}\left[\begin{array}{l} \left\{\frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} \\ +\left\{\bar{y}_r + \hat{b}\left\{\left\{\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} - \bar{x}_r\right\}\right. \end{array}\right]$$

$$\bar{y}_{MI_4} = \frac{1}{2}\left[\tau_1 + \tau_3\right]$$

$$(2.11)\qquad \bar{y}_{MI_4} = \frac{1}{2}\left[\begin{array}{l} \left\{\bar{y}_r + \hat{b}\left(\bar{x}_n - \bar{x}_r\right) exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} \\ +\left\{\bar{y}_r + \hat{b}\left\{\left\{\bar{x}_n exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)\right\} - \bar{x}_r\right\}\right. \end{array}\right]$$

3. Bias and mean square errors of the proposed estimators $\tau_1, \tau_2, \tau_3, \bar{y}_{MI_1}, \bar{y}_{MI_2}, \bar{y}_{MI_3}$ and $\bar{y}_{MI_4}$

Under the suggested method of imputation the estimators $\tau_1, \tau_2, \tau_3, \bar{y}_{MI_1}, \bar{y}_{MI_2}$, $\bar{y}_{MI_3}$ and $\bar{y}_{MI_4}$defined in equations (2.3), (2.5), (2.7) and (2.8)-(2.11) are biased estimators of $\bar{Y}$. Since, we have considered the MCAR response mechanism, therefore, the bias and mean square errors of the proposed estimators are derived up to the first order of approximations using the following transformations:
$\bar{y}_r = \bar{Y}\left(1+e_1\right), \bar{x}_n = \bar{X}\left(1+e_2\right), \bar{x}_r = \bar{X}\left(1+e_3\right), s_{yx}(r) = S_{yx}\left(1+e_4\right)$,
$s_x^2(r) = S_x^2\left(1+e_5\right)$ such that $E(e_i) = 0$ and $|e_i| < 1$ for i=1,2,...,5.
Under the above transformation, the estimators τ_1, τ_2 and τ_3 take the following forms:

$$(3.1)\qquad \tau_1 = \left[\begin{array}{l} \left\{\bar{Y}\left(1+e_1\right) + \beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\left(e_2 - e_3\right)\right\} \\ exp\left\{\text{-}\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\} \end{array}\right]$$

$$(3.2)\qquad \tau_2 = \left[\left\{\bar{Y}\left(1+e_1\right)\left(1+e_2\right)\left(1+e_3\right)^{-1}\right\} exp\left\{\text{-}\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right]$$

$$(3.3)\qquad \tau_3 = \left[\begin{array}{l} \left\{\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\right\} \\ \left\{\left\{\left(1+e_2\right)exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right\}-\left(1+e_3\right)\right\} \end{array}\right]$$

The bias and the mean square errors up to the first order of approximations of the proposed estimators $\tau_1, \tau_2, \tau_3, \bar{y}_{MI_1}, \bar{y}_{MI_2}, \bar{y}_{MI_3}$ and $\bar{y}_{MI_4}$ are derived in the following theorems:

3.1. Theorem. *The bias of the estimators $\tau_1, \tau_2, \tau_3, \bar{y}_{MI_1}, \bar{y}_{MI_2}, \bar{y}_{MI_3}$ and $\bar{y}_{MI_4}$ are given by*

$$(3.4)\qquad B\left(\tau_1\right) = \left[\begin{array}{l} \bar{Y}\left\{\left(\frac{1}{r}-\frac{1}{N}\right)\frac{1}{2}\left(\frac{3}{4}\frac{\mu_{200}}{\bar{X}^2}-\frac{\mu_{110}}{\bar{X}\bar{Y}}\right)\right\} \\ +\left\{\left(\frac{1}{r}-\frac{1}{n}\right)\beta_{yx}\left(\frac{1}{2}\frac{\mu_{200}}{\bar{X}}+\frac{\mu_{300}}{\mu_{200}}-\frac{\mu_{210}}{\mu_{110}}\right)\right\} \end{array}\right]$$

$$(3.5)\qquad B\left(\tau_2\right) = \bar{Y}\left[\begin{array}{l} \left\{\left(\frac{1}{n}-\frac{1}{N}\right)\left(\rho_{yx}C_yC_x-\frac{3}{2}C_x^2\right)\right\} \\ +\left\{\left(\frac{1}{r}-\frac{1}{N}\right)\frac{1}{2}\left(\frac{15}{4}C_x^2-3\rho_{yx}C_yC_x\right)\right\} \end{array}\right]$$

$$(3.6)\qquad B\left(\tau_3\right) = \beta_{yx}\left[\begin{array}{l} \left\{\left(\frac{1}{r}-\frac{1}{N}\right)\frac{3}{2}\left(\frac{1}{4}\frac{\mu_{200}}{\bar{X}}+\frac{\mu_{300}}{\mu_{200}}-\frac{\mu_{210}}{\mu_{110}}\right)\right\} \\ +\left\{\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{\mu_{210}}{\mu_{110}}-\frac{\mu_{300}}{\mu_{200}}-\frac{1}{2}\frac{\mu_{200}}{\bar{X}}\right)\right\} \end{array}\right]$$

$$(3.7)\qquad B(\bar{y}_{MI_1}) = \frac{1}{3}\left\{B(\tau_1)+B(\tau_2)+B(\tau_3)\right\}$$

$$(3.8)\qquad B(\bar{y}_{MI_2}) = \frac{1}{2}\left\{B(\tau_1)+B(\tau_2)\right\}$$

$$(3.9)\qquad B(\bar{y}_{MI_3}) = \frac{1}{2}\left\{B(\tau_2)+B(\tau_3)\right\}$$

$$(3.10)\qquad B(\bar{y}_{MI_4}) = \frac{1}{2}\left\{B(\tau_1)+B(\tau_3)\right\}$$

where $\mu_{rst} = E\left[\left(x_i-\bar{X}\right)^r\left(y_i-\bar{Y}\right)^s\left(z_i-\bar{Z}\right)^t\right]; (r,s,t) \geq 0$ *are integers.*

$C_y^2 = \frac{S_y^2}{\bar{Y}^2}, C_x^2 = \frac{S_x^2}{\bar{X}^2}, \rho_{yx} = \frac{S_{yx}}{S_yS_x}$, S_y^2, S_x^2 *and* S_{yx} *have their usual meanings.*

Proof. The bias of the estimators τ_1, τ_2 and τ_3 are derived as

$B\left(\tau_1\right) = E\left[\tau_1-\bar{Y}\right]$

$$(3.11)\qquad = E\left[\left[\begin{array}{l} \left\{\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\left(e_2-e_3\right)\right\} \\ exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\} \end{array}\right]-\bar{Y}\right]$$

$B\left(\tau_2\right) = E\left[\tau_2-\bar{Y}\right]$

$$(3.12)\qquad = E\left[\left[\left\{\bar{Y}\left(1+e_1\right)\left(1+e_2\right)\left(1+e_3\right)^{-1}\right\}exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right]-\bar{Y}\right]$$

$B\left(\tau_3\right) = E\left[\tau_3-\bar{Y}\right]$

$$(3.13)\qquad = E\left[\left[\begin{array}{l} \left\{\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1} \right. \\ \left\{\left\{\left(1+e_2\right)exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right\}-\left(1+e_3\right)\right\} \end{array}\right]-\bar{Y}\right]$$

Now, expanding the right hand side of the equations (3.11) - (3.13) binomially and exponentially, taking expectations and retaining the terms up to first order of approximations, we get the expressions of the bias of the estimators τ_1, τ_2 and τ_3 as derived in equations (3.4) - (3.6).

The bias of the estimators $\bar{y}_{MI_1}, \bar{y}_{MI_2}, \bar{y}_{MI_3}$ and $\bar{y}_{MI_4}$ are derived as

$$B\left(\bar{y}_{MI_1}\right) = E\left[\bar{y}_{MI_1} - \bar{Y}\right]$$

$$= E\left[\left\{\frac{1}{3}\left\{\tau_1 + \tau_2 + \tau_3\right\}\right\} - \bar{Y}\right] = \frac{1}{3}E\left[\left(\tau_1 - \bar{Y}\right) + \left(\tau_2 - \bar{Y}\right) + \left(\tau_3 - \bar{Y}\right)\right]$$

$$= \frac{1}{3}\left[E\left(\tau_1 - \bar{Y}\right) + E\left(\tau_2 - \bar{Y}\right) + E\left(\tau_3 - \bar{Y}\right)\right]$$

$$B\left(\bar{y}_{MI_1}\right) = \frac{1}{3}\left\{B(\tau_1) + B(\tau_2) + B(\tau_3)\right\} \tag{3.14}$$

$$B\left(\bar{y}_{MI_2}\right) = E\left[\bar{y}_{MI_2} - \bar{Y}\right]$$

$$= E\left[\left\{\frac{1}{2}\left\{\tau_1 + \tau_2\right\} - \bar{Y}\right] = \frac{1}{2}E\left[\left(\tau_1 - \bar{Y}\right) + \left(\tau_2 - \bar{Y}\right)\right]$$

$$= \frac{1}{2}\left[E\left(\tau_1 - \bar{Y}\right) + E\left(\tau_2 - \bar{Y}\right)\right]$$

$$B\left(\bar{y}_{MI_2}\right) = \frac{1}{2}\left\{B(\tau_1) + B(\tau_2)\right\} \tag{3.15}$$

$$B\left(\bar{y}_{MI_3}\right) = E\left[\bar{y}_{MI_3} - \bar{Y}\right]$$

$$= E\left[\left\{\frac{1}{2}\left\{\tau_2 + \tau_3\right\}\right\} - \bar{Y}\right] = \frac{1}{2}E\left[\left(\tau_2 - \bar{Y}\right) + \left(\tau_3 - \bar{Y}\right)\right]$$

$$= \frac{1}{2}\left[E\left(\tau_2 - \bar{Y}\right) + E\left(\tau_3 - \bar{Y}\right)\right]$$

$$B\left(\bar{y}_{MI_3}\right) = \frac{1}{2}\left\{B(\tau_2) + B(\tau_3)\right\} \tag{3.16}$$

$$B\left(\bar{y}_{MI_4}\right) = E\left[\bar{y}_{MI_4} - \bar{Y}\right]$$

$$= E\left[\left\{\frac{1}{2}\left\{\tau_1 + \tau_3\right\}\right\} - \bar{Y}\right] = \frac{1}{2}E\left[\left(\tau_1 - \bar{Y}\right) + \left(\tau_3 - \bar{Y}\right)\right]$$

$$= \frac{1}{2}\left[E\left(\tau_1 - \bar{Y}\right) + E\left(\tau_3 - \bar{Y}\right)\right]$$

$$B\left(\bar{y}_{MI_4}\right) = \frac{1}{2}\left\{B(\tau_1) + B(\tau_3)\right\} \tag{3.17}$$

where $B\left(\tau_1\right) = E\left[\tau_1 - \bar{Y}\right]$, B$(\tau_2) = E\left[\tau_2 - \bar{Y}\right]$ and $B\left(\tau_3\right) = E\left[\tau_3 - \bar{Y}\right]$

□

3.2. Theorem. *The mean square errors of the estimators* $\tau_1, \tau_2, \tau_3, \bar{y}_{MI_1}, \bar{y}_{MI_2}, \bar{y}_{MI_3}$ *and* $\bar{y}_{MI_4}$ *are given by*

$$M(\tau_1) = \bar{Y}^2 \left[\begin{array}{l} \left(\frac{1}{r} - \frac{1}{N}\right)\left\{C_y^2 + \frac{1}{4}C_x^2 - \rho_{yx}C_yC_x\right\} \\ + \left(\frac{1}{r} - \frac{1}{n}\right)\rho_{yx}C_yC_x\left\{C_x - \rho_{yx}C_y\right\} \end{array} \right] \tag{3.18}$$

$$M(\tau_2) = \bar{Y}^2 \left[\begin{array}{l} \left(\frac{1}{r} - \frac{1}{N}\right)\left\{C_y^2 + \frac{9}{4}C_x^2 - 3\rho_{yx}C_yC_x\right\} \\ +2\left(\frac{1}{n} - \frac{1}{N}\right)\left\{\rho_{yx}C_yC_x - C_x^2\right\} \end{array} \right] \tag{3.19}$$

$$M(\tau_3) = \bar{Y}^2C_y^2\left(\frac{1}{r} - \frac{1}{N}\right)\left[1 - \frac{3}{4}\rho_{yx}^2\right] \tag{3.20}$$

$$M(\bar{y}_{MI_1}) = \left[\begin{array}{l} \frac{1}{9}\left[M(\tau_1) + M(\tau_2) + M(\tau_3)\right] \\ +2\left\{C\left(\tau_1, \tau_2\right) \ + C\left(\tau_1, \tau_3\right) \ + C\left(\tau_2, \tau_3\right)\right\} \end{array} \right] \tag{3.21}$$

$$M(\bar{y}_{MI_2}) = \frac{1}{4}\left[M(\tau_1) + M(\tau_2) + 2C\left(\tau_1, \tau_2\right)\right] \tag{3.22}$$

$$M(\bar{y}_{MI_3}) = \frac{1}{4}\left[M(\tau_2) + M(\tau_3) \ + 2C\left(\tau_2, \tau_3\right)\right] \tag{3.23}$$

$$M(\bar{y}_{MI_4}) = \frac{1}{4}\left[M(\tau_1) + M(\tau_3) \ + 2C\left(\tau_1, \tau_3\right)\right] \tag{3.24}$$

where

$$C\left(\tau_1, \tau_2\right) = \bar{Y}^2 \left[\begin{array}{l} \left(\frac{1}{r} - \frac{1}{N}\right)\left(C_y^2 - \frac{1}{4}C_x^2 - \rho_{yx}C_yC_x\right) \\ + \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - \rho_{yx}^2C_y^2\right) \end{array} \right] \tag{3.25}$$

$$C\left(\tau_1, \tau_3\right) = \bar{Y}^2 \left[\begin{array}{l} \left(\frac{1}{r} - \frac{1}{N}\right)\left(C_y^2 - \frac{1}{4}\rho_{yx}C_yC_x - \frac{1}{2}\rho_{yx}^2C_y^2\right) \\ + \left(\frac{1}{r} - \frac{1}{n}\right)\frac{1}{2}\left(\rho_{yx}C_yC_x - \rho_{yx}^2C_y^2\right) \end{array} \right] \tag{3.26}$$

$$C\left(\tau_2, \tau_3\right) = \bar{Y}^2 \left[\begin{array}{l} \left(\frac{1}{r} - \frac{1}{N}\right)\left(C_y^2 - \frac{1}{4}\rho_{yx}C_yC_x - \frac{1}{2}\rho_{yx}^2C_y^2\right) \\ + \left(\frac{1}{r} - \frac{1}{n}\right)\left(\rho_{yx}C_yC_x - \rho_{yx}^2C_y^2\right) \end{array} \right] \tag{3.27}$$

Proof. The mean square errors of the estimators τ_1, τ_2 and τ_3 are derived as

$M\left(\tau_1\right) = E\left[\tau_1 - \bar{Y}\right]^2$

$$= E\left[\left[\begin{array}{l} \left\{\bar{Y}\left(1 + e_1\right) + \beta_{yx}\bar{X}\left(1 + e_4\right)\left(1 + e_5\right)^{-1}\left(e_2 - e_3\right)\right\} \\ exp\left\{-\frac{e_3}{2}\left(1 + \frac{e_3}{2}\right)^{-1}\right\} \end{array} \right] - \bar{Y}\right]^2 \tag{3.28}$$

$M\left(\tau_2\right) = E\left[\tau_2 - \bar{Y}\right]^2$

$$= E\left[\left[\left\{\bar{Y}\left(1 + e_1\right)\left(1 + e_2\right)\left(1 + e_3\right)^{-1}\right\} exp\left\{-\frac{e_3}{2}\left(1 + \frac{e_3}{2}\right)^{-1}\right\}\right] - \bar{Y}\right]^2 \tag{3.29}$$

$M\left(\tau_3\right) = E\left[\tau_3 - \bar{Y}\right]^2$

$$= E\left[\left[\begin{array}{l} \left\{\bar{Y}\left(1 + e_1\right) + \beta_{yx}\bar{X}\left(1 + e_4\right)\left(1 + e_5\right)^{-1} \right. \\ \left. \left\{\left\{\left(1 + e_2\right) exp\left\{-\frac{e_3}{2}\left(1 + \frac{e_3}{2}\right)^{-1}\right\}\right\} - \left(1 + e_3\right)\right\} \right. \end{array} \right] - \bar{Y}\right]^2 \tag{3.30}$$

Now, expanding the right hand side of the equations (3.28) - (3.30) binomially and exponentially, taking expectations and retaining the terms up to first order of approximations, we get the expressions of the mean square errors of the estimators τ_1, τ_2 and τ_3 as derived in equations (3.18) - (3.20).

The mean square errors of the estimators $\bar{y}_{MI_1}, \bar{y}_{MI_2}, \bar{y}_{MI_3}$ and $\bar{y}_{MI_4}$ are derived as

$$M\left(\bar{y}_{MI_1}\right) = E\left[\bar{y}_{MI_1} - \bar{Y}\right]^2$$

$$= E\left[\left\{\frac{1}{3}\left\{\tau_1 + \tau_2 + \tau_3\right\}\right\} - \bar{Y}\right]^2 = E\left[\frac{1}{3}\left(\tau_1 - \bar{Y}\right) + \frac{1}{3}\left(\tau_2 - \bar{Y}\right) + \frac{1}{3}\left(\tau_3 - \bar{Y}\right)\right]^2$$

$$M\left(\bar{y}_{MI_1}\right) = \begin{bmatrix} \frac{1}{9}\left\{E\left(\tau_1 - \bar{Y}\right)^2 + E\left(\tau_2 - \bar{Y}\right)^2 + E\left(\tau_3 - \bar{Y}\right)^2\right\} \\ \frac{2}{9}\left[E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_2 - \bar{Y}\right)\right] + E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_3 - \bar{Y}\right)\right]\right] \\ +\frac{2}{9}\left[E\left[\left(\tau_2 - \bar{Y}\right)\left(\tau_3 - \bar{Y}\right)\right]\right] \end{bmatrix} \tag{3.31}$$

$$M\left(\bar{y}_{MI_2}\right) = E\left[\bar{y}_{MI_2} - \bar{Y}\right]^2$$

$$= E\left[\left\{\frac{1}{2}\left\{\tau_1 + \tau_2\right\}\right\} - \bar{Y}\right]^2 = E\left[\frac{1}{2}\left(\tau_1 - \bar{Y}\right) + \frac{1}{2}\left(\tau_2 - \bar{Y}\right)\right]^2$$

$$= \left[\frac{1}{4}E\left(\tau_1 - \bar{Y}\right)^2 + \frac{1}{4}E\left(\tau_2 - \bar{Y}\right)^2 + \frac{1}{2}E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_2 - \bar{Y}\right)\right]\right]$$

$$M(\bar{y}_{MI_2}) = \frac{1}{4}\left[M(\tau_1) + M(\tau_2) + 2C\left(\tau_1, \tau_2\right)\right] \tag{3.32}$$

$$M\left(\bar{y}_{MI_3}\right) = E\left[\bar{y}_{MI_3} - \bar{Y}\right]^2$$

$$= E\left[\left\{\frac{1}{2}\left\{\tau_2 + \tau_3\right\}\right\} - \bar{Y}\right]^2 = E\left[\frac{1}{2}\left(\tau_2 - \bar{Y}\right) + \frac{1}{2}\left(\tau_3 - \bar{Y}\right)\right]^2$$

$$= \left[\frac{1}{4}E\left(\tau_2 - \bar{Y}\right)^2 + \frac{1}{4}E\left(\tau_3 - \bar{Y}\right)^2 + \frac{1}{2}E\left[\left(\tau_2 - \bar{Y}\right)\left(\tau_3 - \bar{Y}\right)\right]\right]$$

$$M(\bar{y}_{MI_3}) = \frac{1}{4}\left[M(\tau_2) + M(\tau_3) \ + 2C\left(\tau_2, \tau_3\right)\right] \tag{3.33}$$

$$M\left(\bar{y}_{MI_4}\right) = E\left[\bar{y}_{MI_4} - \bar{Y}\right]^2$$

$$= E\left[\left\{\frac{1}{2}\left\{\tau_1 + \tau_3\right\}\right\} - \bar{Y}\right]^2 = E\left[\frac{1}{2}\left(\tau_1 - \bar{Y}\right) + \frac{1}{2}\left(\tau_3 - \bar{Y}\right)\right]^2$$

$$= \left[\frac{1}{4}E\left(\tau_1 - \bar{Y}\right)^2 + \frac{1}{4}E\left(\tau_3 - \bar{Y}\right)^2 + \frac{1}{2}E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_3 - \bar{Y}\right)\right]\right]$$

$$M(\bar{y}_{MI_4}) = \frac{1}{4}\left[M(\tau_1) + M(\tau_3) \ + 2C\left(\tau_1, \tau_3\right)\right] \tag{3.34}$$

where $M\left(\tau_1\right) = E\left[\tau_1 - \bar{Y}\right]^2$, M$(\tau_2) = E\left[\tau_2 - \bar{Y}\right]^2$, M$(\tau_3) = E\left[\tau_3 - \bar{Y}\right]^2$, $C\left(\tau_1, \tau_2\right) = E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_2 - \bar{Y}\right)\right]$, $C\left(\tau_1, \tau_3\right) = E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_3 - \bar{Y}\right)\right]$ and $C\left(\tau_2, \tau_3\right) = E\left[\left(\tau_2 - \bar{Y}\right)\left(\tau_3 - \bar{Y}\right)\right]$ The expressions of $C\left(\tau_1, \tau_2\right), C\left(\tau_1, \tau_3\right)$ and $C\left(\tau_2, \tau_3\right)$ are derived as

$$C\left(\tau_1, \tau_2\right) = E\left[\left(\tau_1 - \bar{Y}\right)\left(\tau_2 - \bar{Y}\right)\right]$$

$$(3.35)\quad = E\left[\begin{array}{l} \left[\left\{\left(\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\left(e_2-e_3\right)\right)\right.\right. \\ \left.\left(exp\left(-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right)\right)\right\}-\bar{Y}] \\ \left[\left[\left\{\bar{Y}\left(1+e_1\right)\left(1+e_2\right)\left(1+e_3\right)^{-1}\right\}exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right]-\bar{Y}\right] \end{array}\right]$$

$$C\left(\tau_1,\tau_3\right)=E\left[\left(\tau_1-\bar{Y}\right)\left(\tau_3-\bar{Y}\right)\right]$$

$$(3.36)\quad = E\left[\begin{array}{l} \left[\left\{\left(\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\left(e_2-e_3\right)\right)\right.\right. \\ \left.\left(exp\left(-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right)\right)\right\}-\bar{Y}] \\ \left[\left\{\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\right.\right. \\ \left\{\left\{\left(1+e_2\right)exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right\}-\left(1+e_3\right)\right\}\}-\bar{Y}] \end{array}\right]$$

$$C\left(\tau_2,\tau_3\right)=E\left[\left(\tau_2-\bar{Y}\right)\left(\tau_3-\bar{Y}\right)\right]$$

$$(3.37)\quad = E\left[\begin{array}{l} \left[\left[\left\{\bar{Y}\left(1+e_1\right)\left(1+e_2\right)\left(1+e_3\right)^{-1}\right\}exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right]-\bar{Y}\right] \\ \left[\left\{\bar{Y}\left(1+e_1\right)+\beta_{yx}\bar{X}\left(1+e_4\right)\left(1+e_5\right)^{-1}\right.\right. \\ \left\{\left\{\left(1+e_2\right)exp\left\{-\frac{e_3}{2}\left(1+\frac{e_3}{2}\right)^{-1}\right\}\right\}-\left(1+e_3\right)\right\}\}-\bar{Y}] \end{array}\right]$$

Now, expanding the right hand side of the equations (3.35)− (3.37) binomially and exponentially, taking expectations and retaining the terms up to the first order of approximations, we get the expressions of the $C\left(\tau_1,\tau_2\right), C\left(\tau_1,\tau_3\right)$ and $C\left(\tau_2,\tau_3\right)$ as derived in equations (3.25) - (3.27). □

4. Some well-known methods of single imputation and resultant estimators

Following are the list of some existing methods of imputation and their resultant estimators which are often practiced in survey sampling.

4.1. Mean method of imputation. The data produced under mean method of imputation is described as

$$(4.1)\qquad y_{.i}=\begin{cases} y_i & if\quad i\in R \\ y_r & if\quad i\in R^c \end{cases}$$

Under the method of imputation discussed in equation (4.1), the point estimator of the population mean $\bar{Y}$ is derived as

$$(4.2)\qquad \bar{y}_M=\frac{1}{n}\sum_{i=1}^{n}y_{.i}=\frac{1}{n}\left[\sum_{i\in R}y_{.i}+\sum_{i\in R^c}y_{.i}\right]=\bar{y}_r$$

which is simplified as

The variance of the estimator $\bar{y}_M$ given in equation (4.2) is obtained under MCAR response mechanism and is given as

$$V\left(\bar{y}_M\right)=\left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}^2C_y^2$$

4.2. Ratio method of imputation. The ratio method of imputation is applied with the help of information obtained on an auxiliary variable x and consequently the data generated is described as

$$y_{.i} = \begin{cases} y_i & if \quad i \in R \\ \hat{b}_r x_i & if \quad i \in R^c \end{cases} \tag{4.3}$$

where $\hat{b}_r = \frac{\sum_{i\in R} y_i}{\sum_{i\in R} x_.} = \frac{\bar{y}_r}{\bar{x}_r}$

Under the method of imputation discussed in equation (4.3), the point estimator of population mean $\bar{Y}$ is derived as

$$\bar{y}_{RAT} = \frac{1}{n}\sum_{i=1}^{n} y_{.i} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \tag{4.4}$$

The bias and mean square error of the estimator $\bar{y}_{RAT}$ are obtained under MCAR response mechanism up to first order of approximations and given as

$$B\left(\bar{y}_{RAT}\right) = \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y} \left(C_x^2 - \rho_{yx} C_y C_x\right) \tag{4.5}$$

$$M\left(\bar{y}_{RAT}\right) = \bar{Y}^2 \left[\left(\frac{1}{r} - \frac{1}{n}\right) C_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right) \left(C_x^2 - \rho_{yx} C_y C_x\right)\right] \tag{4.6}$$

4.3. Regression method of imputation. The data generated by regression method of imputation is given as

$$y_{.i} = \begin{cases} y_i & if \quad i \in R \\ \hat{y}_i & if \quad i \in R^c \end{cases} \tag{4.7}$$

where

$$\hat{y}_i = \hat{a} + \hat{b}_{re} x_i, \hat{a} = \bar{y}_r - \hat{b}\bar{x}_r \text{ and } \hat{b}_{re} = \frac{S_{yx}(r)}{S_x^2(r)}$$

Under the method of imputation discussed in equation (4.5), the point estimator of population mean $\bar{Y}$ is derived as

$$\bar{y}_{REG} = \frac{1}{n}\sum_{i=1}^{n} y_{.i} = \bar{y}_r + \hat{b}_{re}\left(\bar{x}_n - \bar{x}_r\right) \tag{4.8}$$

The bias and mean square error of the estimator $\bar{y}_{REG}$ are obtained under MCAR response mechanism up to first order of approximations and given as

$$B\left(\bar{y}_{REG}\right) = \frac{\rho_{yx} C_y}{C_x \bar{X}} \left(\frac{1}{r} - \frac{1}{n}\right) \bar{Y} \left(\frac{\mu_{300}}{\mu_{200}} - \frac{\mu_{210}}{\mu_{110}}\right) \tag{4.9}$$

$$M\left(\bar{y}_{REG}\right) = \bar{Y}^2 C_y^2 \left[\left(\frac{1}{r} - \frac{1}{n}\right) - \left(\frac{1}{r} - \frac{1}{n}\right) \rho_{yx}^2\right] \tag{4.10}$$

5. Empirical study

In this section, we demonstrate the performances of the proposed imputation methods over mean, ratio and regression methods of imputation. To access the performances of the proposed methods, empirical studies are carried out on seventeen natural populations chosen from various survey literatures related to life sciences, agricultural and socio-economic characters. The details of the populations are provided in this section. The methodology of empirical study is as follows; from a finite population of size N a sample of size n is drawn under SRSWOR sampling scheme. The first m samples were selected from the all possible ${}^{N}C_n$ samples. First we drop (n-r) units randomly from each sample corresponding to the study variable y and imputed values are derived with six methods of imputations namely (i) Mean method of imputation (ii) Ratio method of imputation (iii) Regression method of imputation (iv) Suggested single imputations methods (v) Suggested multiple imputations methods

The percent relative efficiencies of the proposed single imputation methods with respect to the mean, ratio and regression methods of imputation are given as

$$PRE_1 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_M)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_1)_s - \bar{Y}\right]^2}\times 100, PRE_2 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_1)_s - \bar{Y}\right]^2}\times 100,$$

$$PRE_3 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_1)_s - \bar{Y}\right]^2}\times 100, PRE_4 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_M)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}\times 100,$$

$$PRE_5 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}\times 100, PRE_6 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}\times 100,$$

$$PRE_7 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_M)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_3)_s - \bar{Y}\right]^2}\times 100, PRE_8 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_3)_s - \bar{Y}\right]^2}\times 100$$

and $PRE_9 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\tau_3)_s - \bar{Y}\right]^2}\times 100$

The percent relative efficiencies of the proposed multiple imputations methods with respect to the mean, ratio, regression and proposed single imputation methods are given as

$$E_1 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_M)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_1})_s - \bar{Y}\right]^2}\times 100, E_2 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_1})_s - \bar{Y}\right]^2}\times 100,$$

$$E_3 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_1})_s - \bar{Y}\right]^2}\times 100, E_4 = \frac{\sum_{s=1}^{m}\left[(\tau_1)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_1})_s - \bar{Y}\right]^2}\times 100,$$

$$E_5 = \frac{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_1})_s - \bar{Y}\right]^2}\times 100, E_6 = \frac{\sum_{s=1}^{m}\left[(\tau_3)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_1})_s - \bar{Y}\right]^2}\times 100,$$

$$E_7 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_M)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_2})_s - \bar{Y}\right]^2}\times 100, E_8 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_2})_s - \bar{Y}\right]^2}\times 100,$$

$$E_9 = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_2})_s - \bar{Y}\right]^2} \times 100, E_{10} = \frac{\sum_{s=1}^{m}\left[(\tau_1)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_2})_s - \bar{Y}\right]^2} \times 100,$$

$$E_{11} = \frac{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_2})_s - \bar{Y}\right]^2} \times 100, E_{12} = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{M})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_3})_s - \bar{Y}\right]^2} \times 100,$$

$$E_{13} = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_3})_s - \bar{Y}\right]^2} \times 100, E_{14} = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_3})_s - \bar{Y}\right]^2} \times 100,$$

$$E_{15} = \frac{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_3})_s - \bar{Y}\right]^2} \times 100, E_{16} = \frac{\sum_{s=1}^{m}\left[(\tau_3)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_3})_s - \bar{Y}\right]^2} \times 100,$$

$$E_{17} = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{M})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_4})_s - \bar{Y}\right]^2} \times 100, E_{18} = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{RAT})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_4})_s - \bar{Y}\right]^2} \times 100,$$

$$E_{19} = \frac{\sum_{s=1}^{m}\left[(\bar{y}_{REG})_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_4})_s - \bar{Y}\right]^2} \times 100, E_{20} = \frac{\sum_{s=1}^{m}\left[(\tau_1)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_4})_s - \bar{Y}\right]^2} \times 100,$$

and $E_{21} = \frac{\sum_{s=1}^{m}\left[(\tau_2)_s - \bar{Y}\right]^2}{\sum_{s=1}^{m}\left[(\bar{y}_{MI_4})_s - \bar{Y}\right]^2} \times 100,$

The percent relative efficiencies are computed for seventeen natural populations as described below and presented in Tables 1-7.

Population I [Source: [12]] (Page No. 399)
Y: Area under wheat in 1964
X: Area under wheat in 1963
N = 34, n = 7, r = 5, $\rho_{yx} = 0.9800867$.

Population II [Source: [3]] (Page No. 58)
Y: Head length of second son.
X: Head length of first son.
$N = 25, n = 7, r = 5, \rho_{yx} = 0.7107518$.

Population III [Source: [5]] (Page No. 182)
Y: Number of placebo children.
X: Number of paralytic polio cases in the placebo group.
$N = 34, n = 7, r = 5, \rho_{yx} = 0.7328235$.

Population IV [Source: [8]] (Page No. 682)
Y: No. of hhś on ith block.
X: Eye estimate of no. of hhś on ith block
$N = 20, n = 7, r = 5, \rho_{yx} = 0.8662052$.

Population V [Source: [24]] (Page No. 349)
Y: Volume.
X: Diameter
$N = 31, n = 7, r = 5, \rho_{yx} = 0.9671194$.

Population VI [Source: [5]] (Page No. 182)
Y: Number of placebo children.
X: Number of paralytic polio cases in the not inoculated group.
$N = 34, n = 7, r = 5, \rho_{yx} = 0.6426412$.

Population VII [Source: [12]] (Page No. 399)

Y: Area under wheat in 1964
X: Cultivated area in 1961
$N = 34, n = 7, r = 5, \rho_{yx} = 0.9042627$.
Population VIII [Source: [3]] (Page No. 58)
Y: Head length of second son.
X: Head breadth of first son.
$N = 34, n = 7, r = 5, \rho_{yx} = 0.6931573$.
Population IX [Source: [5]] (Page No. 34)
Y: Food cost of family
X: Size of family
$N = 33, n = 7, r = 5, \rho_{yx} = 0.432738$.
Population X [Source: [7]] (Page No. 180)
Y: Sepal width of Iris setosa
X: Sepal length of Iris setosa
$N = 35, n = 7, r = 5, \rho_{yx} = 0.6315548$.
Population XI [Source: [6]] (Page No. 154)
Y: Average salary (in dollars) U. S.
X: Per pupil spending (in dollars) U. S.
$N = 26, n = 7, r = 5, \rho_{yx} = 0.8096703$.
Population XII [Source: [6]] (Page No. 274)
Y: Saving (in billions of dollars) U. S. (1970-1995).
X: Personal disposable income (in billions of dollars) U. S. (1970-1995).
$N = 26, n = 7, r = 5, \rho_{yx} = 0.8759079$.
Population XIII [Source: [6]] (Page No. 460)
Y: Index of real compensation per hour, business sector of U. S. (1959-1998).
X: Index of output per hour, business sector of U. S. (1959-1998).
$N = 30, n = 7, r = 5, \rho_{yx} = 0.9910549$.
Population XIV [Source: [6]] (Page No. 710)
Y: Investment in fixed plant and equipment in manufacturing (in billions of dollars) of U. S. (1970-1991).
X: Manufacturing sales (in billions of dollars) seasonally adjusted of U. S. (1970-1991).
$N = 22, n = 7, r = 5, \rho_{yx} = 0.9903192$.
Population XV [Source: [23]] (Page No. 166)
Y: Number of banana bunches.
X: Number of banana pits.
$N = 20, n = 7, r = 5, \rho_{yx} = 0.9800867$.
Population XVI [Source: [24]] (Page No. 349)
Y: Volume.
Z: Height
$N = 31, n = 7, r = 5, \rho_{yx} = 0.5982497$.
Population XVII [Source: [5]] (Page No. 32)
Y: Food cost of family
X: Income of family
$N = 33, n = 7, r = 5, \rho_{yx} = 0.2521603$.

Table 1: Percent relative efficiencies of the estimator τ_1 with respect to mean, ratio and regression method of imputation

Population Source	PRE_1	PRE_2	PRE_3
Population I	651.309	316.1384	323.7037
Population II	157.1894	126.3349	124.532
Population III	223.1392	162.9272	194.0364
Population IV	294.5976	188.9788	186.6463
Population V	164.7055	154.3641	158.9833
Population VI	200.7349	166.7052	181.3413
Population VII	284.5805	182.3409	178.0122
Population VIII	241.128	170.1591	155.7499
Population IX	146.6306	133.258	110.9385
Population X	100.5127	106.255	101.159
Population XI	182.2423	144.3668	142.4705
Population XII	264.9797	189.8048	184.0865
Population XIII	2139.517	735.6239	925.6935
Population XIV	287.5206	237.6237	237.7244
Population XV	236.8863	169.4697	172.0994

Table 2: Percent relative efficiencies of the estimator τ_2 with respect to mean, ratio and regression method of imputation

Population Source	PRE_1	PRE_2	PRE_3
Population I	609.8675	296.0231	303.1071
Population II	125.24594	100.6827	100.24594
Population III	177.9285	129.9161	154.7222
Population IV	248.101	147.7621	150.0241
Population V	301.875	282.9211	291.3873
Population VI	143.1064	118.8463	129.3873
Population VII	245.6476	157.3952	153.6587
Population VIII	181.8826	127.9263	117.0935
Population IX	116.9035	111.8727	106.1338
Population X	145.7711	115.4754	113.9586
Population XI	163.0738	142.7995	138.4974
Population XII	193.4761	198.1857	205.0121
Population XIII	3647.527	1254.118	1578.156
Population XIV	316.3238	261.4263	261.5392
Population XV	208.6929	149.2999	151.6167

Table 3: Percent relative efficiencies of the estimator τ_3 with respect to mean, ratio and regression method of imputation

Population Source	PRE_1	PRE_2	PRE_3
Population I	746.0278	362.1138	370.7794
Population II	148.3297	119.2142	117.5129
Population III	136.2724	100.50058	118.4991
Population IV	287.3633	184.3382	182.063
Population V	158.7588	148.7907	153.2432
Population VI	121.4111	100.8289	109.6812
Population VII	339.7371	217.6817	212.5141
Population VIII	261.7878	184.1273	168.5353
Population IX	149.2613	135.6488	112.9288
Population X	105.8197	111.8657	106.5001
Population XI	174.4859	138.224	136.4069
Population XII	241.7096	247.5934	256.1216
Population XIII	1264.535	434.813	547.1196
Population XVI	307.6482	254.2538	254.3616
Population XV	236.1414	168.9367	171.5582

Table 4: Percent relative efficiencies of the estimator $\bar{y}_{MI_1}$ with respect to mean, ratio, regression, τ_1, τ_2, and τ_3 method of imputation

Source	E_1	E_2	E_3	E_4	E_5	E_6
Population I	264.907	207.769	204.951	101.495	100.9956	100.4527
Population XVI	161.723	149.986	123.486	111.909	131.157	115.262
Population XVII	108.2	101.727	104.704	108.014	107.128	126.231

Table 5: Percent relative efficiencies of the estimator $\bar{y}_{MI_2}$ with respect to mean, ratio, regression, τ_1, and τ_2 method of imputation

Source	E_7	E_8	E_9	E_{10}	E_{11}
Population I	263.29	206.5075	203.701	100.8759	100.38554
Population I	204.77	134.7842	157.6394	105.1446	108.5116
Population I	192.0412	148.8212	173.3055	106.0514	111.1025
Population I	142.7501	122.8817	144.1724	108.6922	101.2051
Population I	239.2887	173.8419	171.1855	101.0125	101.3311
Population I	107.9156	101.4604	104.4289	107.7306	106.847

Table 6: Percent relative efficiencies of the estimator $\bar{y}_{MI_3}$ with respect to mean, ratio, regression, τ_2, and τ_3 method of imputation

Source	E_{12}	E_{13}	E_{14}	E_{15}	E_{16}
Population I	578.9268	251.5427	240.5896	102.2776	101.6885
Population I	189.3006	124.63	145.77	100.3426	138.7786
Population I	280.08	169.36	166.809	102.6493	112.8907
Population I	143.5103	131.2604	113.5755	110.1633	103.5136
Population I	120.7001	109.198	114.8592	101.4147	109.9245
Population I	314.2731	259.8436	259.7335	102.1552	100.3516
Population I	169.3458	157.0557	129.3067	137.3398	120.6951
Population I	109.7599	103.1943	106.2136	108.6731	128.0516

Table 7: Percent relative efficiencies of the estimator $\bar{y}_{MI_4}$ with respect to mean, ratio, regression, τ_1, and τ_3 method of imputation

Source	E_{17}	E_{18}	E_{19}	E_{20}	E_{21}
Population I	264.0247	207.0764	204.268	101.1567	100.1214
Population I	118.5684	118.6753	128.109	100.5546	101.4809
Population I	152.753	127.9926	137.42	100.3505	101.3829
Population I	155.0982	143.8421	118.4277	107.3255	110.5407

6. Conclusions and recommendations

A close look on Tables 1-7 reveals that the proposed methods of imputations are rewarding in terms of percent relative efficiencies. These findings suggest that the proposed single and multiple methods of imputations described in this paper are highly beneficial in minimizing the negative impact of non-response to a greater extent as compared to the mean, ratio and regression methods of imputation. The survey statisticians may be encouraged for the practical applications of the suggested imputation methods, if non-response is unavoidable in the survey data.

References

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

[2] Al-Omari, A. I., Bouza C.N, Herrera, C. *Imputation methods of missing data for estimating the population mean using simple random sampling with known correlation coefficient*, Quality and Quantity **47** (1), 353-365, 2013.

[3] Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc.,New York. 1958.

[4] Bouza C.N. *Estimation of the population mean with missing observations using product type estimators*, Revista Investigacion Operacional **29** (3), 207-233, 2008.

[5] Cochran, W. G. *Sampling techniques*, New-York, John Wiley and Sons 1977.

[6] Fisher, R. A. *The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics*, Statistics in Transition **7** 179-188, 1936.

[7] Gujrati D. N. *Basic econometrics*, McGraw-Hill higher education 2003.

[8] Horvitz D. G., Thompson D. J. *A generalization of sampling without replacement from a finite universe*, Journal of American Statistical Association **47**, 663-685, 1952.

[9] Kalton, G., Kasprzyk, D. and Santos, R. *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.)*, Academic Press 455-480, 1981.

[10] Lee, H., Rancourt, E. and Sarndal,C. E., *Experiments with variance estimation from survey data with imputed values*, Journal of official Statistics, **10** (3), 231-243, 1994.

[11] Lee, H., Rancourt, E. and Sarndal,C. E. *Variance estimation in the presence of imputed data for the generalized estimation system*, Proceedings. of the American Statistical Association (Social Survey Research Methods Sec.) 384-389, 1995.

[12] Murthy, M. N. *Sampling theory and methods*, Calcutta, India (Statistical Publishing Society) 1967.

[13] Rubin, R. B. *Inference and missing data*, Biometrika 63, 581-592, 1976.

[14] Rubin, R. B. *Formalizing Subjective Notion about the Effect of Non-respondents in Sample Surveys* , Journal of the American Statistical Association 72, 538-543, 1977.

[15] Rubin, R. B. *Multiple imputation for non-response in surveys*, John Wiley, New York 1987.

[16] Sande, I. G. *A personal view of hot deck approach to automatic edit and imputation*, Journal Imputation Procedures Survey Methodology 5, 238-246, 1979.

[17] Singh, G. N. and Karna, J. P. *Some imputation methods to minimize the effect of non response in two-occasion rotation patterns*, Communications in Statistics-Theory and Methods **39** (18), 3264-3281, 2010.

[18] Singh, G. N., Priyanka, K. and Kozak, M. *Use of imputation methods at current occasion in two-occasion rotation patterns.*, Model Assisted Statistical and Applications **3** (2), 99-112, 2008.

[19] Singh, G. N. and Priyanka, K. *): Use of imputation methods in two-occasion successive sampling*, Journal of the Indian Society of Agricultural Statistics **64** (3), 423-438, 2010.

[20] Singh, S. and Horn, S. *Compromised imputation in survey sampling*, Metrika (51), 266-276, 2000.

[21] Singh, S. *A new method of imputation in survey sampling*, Statistics **43** (5), 499-511, 2009.

[22] Singh, S. Deo, B. *Imputation by power transformation*, Statistical Papers (44), 555-579, 2003.

[23] Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok, C. *Sampling theory of surveys with applications*, Iowa State University Press, Ames, Iowa (USA) and Indian Society of Agricultural Statistics, New Delhi (India) 1984.

[24] Wang S. G., Chow S. C. *Advanced Linear Models: Theory and Applications.*, New York, Marcel Dekker, Inc. 1994.