Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110214

An Improved Alternative Method of Imputation for Missing Data in Survey Sampling

G. N. Singh[1], *Ashok. K. Jaiswal*[1,*], *Chandraketu Singh*[2] *and M. Usman*[1]

[1] Indian Institute of Technology (Indian School of Mines), Dhanbad, India - 826004
[2] Department of Statistics, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

Received: 27 Feb. 2020, Revised: 12 May. 2020, Accepted: 26 Jun. 2020
Published online: 1 May 2022

Abstract: In the present paper, a new and improved method of ratio type imputation and corresponding point estimator to estimate the finite population mean is proposed in case of missing data problem. It has been shown that this estimator utilizes the readily available auxiliary information efficiently and gives better results than the ratio and mean methods of imputation; furthermore, its efficiency is also compared with the regression method of imputation and some other imputation methods, discussed in this article, using four real data sets. A simulation study is carried out to verify theoretical outcomes, and suitable recommendations are made.

Keywords: Population mean, imputation, bias, mean squared error (MSE), percent relative efficiency (PRE)

1 Introduction

Nowadays, missing data (non-response) is a common and unavoidable problem in sample surveys. Sample survey experts have recognized for some time that failure to account for the random nature of incompleteness or non-response can ruin the nature of data. There are two types of non-responses occurring in surveys: unit non-response and item non-response. Unit non-response occurs when an eligible sample unit is completely absent whereas item non-response occurs when sampled unit is present in the survey but fails to provide information about some component of a unit in sample survey. Such situations causes missing data problem. To handle this problem, several researchers used imputation method as a process of replacing missing data with substituted values. It is a highly recommended procedure to settle up non-responses in sample survey problems. [1] introduced three concepts: MAR (missing at random), OAR (observed at random) and PD (parameter distinctness). Subsequently, the conceptual difference between MAR and MCAR was discussed by [2]. Several authors [3,4,5,6,7,8,9,10] and[11], assumed MCAR mechanism to suggest several imputation methods and their resultant point estimators for estimation of population mean under non response situations. Motivated with the above works, we suggest an improved estimation procedure of population mean under MCAR mechanism. The behaviors (properties) of the proposed estimation procedure have been examined up to the first order of approximation. The predominance of the proposed estimation procedure over other existing estimators has been shown on three population data sets through empirical studies. Simulation studies are also carried out to verify the theoretical findings and suitable recommendations are made. Consider a finite population $U=(U_1,U_2,...,U_N)$ of size N with population mean $\bar{Y}=N^{-1}\sum_{i=1}^{N}y_i$. To estimate the population mean $\bar{Y}$ of the study variable Y, a sample s of sample size n using simple random sampling without replacement (SRSWOR) method is drawn. Let r be the no. of responding units out of sampled n units. Let R and $\bar{R}$ be the set of responding units and non-responding units respectively. For every unit selected from R, its value y_i is observed whereas for the units belonging to $\bar{R}$, values are missing and imputed values are to be derived. In many cases imputation is done using some quantitative auxiliary variable x. Let x_i be the value of the i^{th} unit of the auxiliary variable x which is

* Corresponding author e-mail: ashok.bhu44@gmail.com

positive for all $i \in s$. Let $y_{.i}$ be the value of the study variable Y and defined by:

$$y_{.i} = \begin{cases} y_i, if\ i \in R \\ \tilde{y}_i, if\ i \in \overline{R} \end{cases} \tag{1}$$

where $\tilde{y}_i$ is the imputed value for the i^{th} non-responding unit. Using above data, we get following form of the general point estimator of population mean $\bar{Y}$:

$$T = \frac{1}{n}\sum_{i=1}^{n} y_{.i} = \frac{1}{n}\left(\sum_{i\in R} y_i + \sum_{i\in \bar{R}} \tilde{y}_i\right) \tag{2}$$

2 Some available imputation methods

2.1 Mean method of imputation

In this method, no auxiliary information is used. The missing values are replaced with the mean of the responding units of the study variable. Hence we get the following data after imputation

$$y_{.i} = \begin{cases} y_i, if\ i \in R \\ \bar{y}_r, if\ i \in \overline{R} \end{cases} \tag{3}$$

So, the resultant point estimator is

$$T_{mean} = \frac{1}{n}\left(\sum_{i\in R} y_i + \sum_{i\in \bar{R}} \bar{y}_r\right) = \frac{1}{r}\sum_{i=1}^{r} y_i = \bar{y}_r \tag{4}$$

It is well known that this is an unbiased estimator. The variance of T_{mean} is obtained as

$$V(T_{mean}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 \tag{5}$$

2.2 Ratio method of imputation

Following the notations of [12], in the case of single Imputation method, if the i^{th} unit requires imputation, the value $\hat{b}x_i$ is imputed, where

$$\hat{b} = \frac{\sum_{i=1}^{r} y_i}{\sum_{i=1}^{r} x_i} \tag{6}$$

Data after imputation becomes

$$y_{.i} = \begin{cases} y_i, if\ i \in R \\ \hat{b}x_i, if\ i \in \overline{R} \end{cases} \tag{7}$$

Hence, the point estimator of population mean $\bar{Y}$ is given as

$$T_{ratio} = \frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n \tag{8}$$

The bias of the point estimator T_{ratio} is given as

$$Bias(T_{ratio}) = \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - \rho_{yx}C_yC_x\right)\bar{Y} \tag{9}$$

The MSE of T_{ratio} is given as

$$MSE(T_{ratio}) = V(T_{mean}) + \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - 2\rho_{yx}C_yC_x\right)\bar{Y}^2 \tag{10}$$

2.3 Regression method of imputation

In this method, the point estimator of population mean $\bar{Y}$ is given as

$$T_{reg} = \bar{y}_r + \hat{b}_{yx}(\bar{x}_n - \bar{x}_r) \tag{11}$$

where $\hat{b}_{yx} = \frac{s_{yx}}{s_x^2}$; $s_{yx} = \frac{1}{r-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)(y_i - \bar{y}_r)$ and $s_x^2 = \frac{1}{r-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2$. The bias of T_{reg} is given by

$$Bias(T_{reg}) = \frac{\rho_{yx}C_y}{C_x\bar{X}}\left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}\left(\frac{\mu_{300}}{\mu_{200}} - \frac{\mu_{210}}{\mu_{110}}\right) \tag{12}$$

where $\mu_{rst} = \sum_{i=1}^{N}(x_i - \bar{X})^r(y_i - \bar{Y})^s(z_i - \bar{Z})^t$. The MSE of T_{reg} is given as

$$MSE(T_{reg}) = \bar{Y}^2C_y^2\left[\left(\frac{1}{r} - \frac{1}{N}\right) - \left(\frac{1}{r} - \frac{1}{n}\right)\rho_{yx}^2\right] \tag{13}$$

2.4 Compromised method of imputation

Singh and Horn (2000) suggested this method, and the data after imputation takes the following form

$$y_{.i} = \begin{cases} \alpha n y_i/r + (1-\alpha)\hat{b}_{yx}x_i, if\ i \in R \\ (1-\alpha)\hat{b}_{yx}x_i, if\ i \in \overline{R} \end{cases} \tag{14}$$

Using above data, we get the point estimator of population mean as following:

$$T_{SH} = \alpha\bar{y}_r + (1-\alpha)\bar{y}_r\frac{\bar{x}_n}{\bar{x}_r} \tag{15}$$

where α is an appropriate constant with optimum value $\alpha^* = 1 - \rho_{yx}\frac{C_y}{C_x}$. The bias of T_{SH} obtained by Singh ad Horn [11] is given by

$$Bias(T_{SH}) = (1-\alpha)\left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}\left(C_x^2 - \rho_{yx}C_yC_x\right) \tag{16}$$

Using α^*, we get the minimum MSE of T_{SH} as

$$MSE_{\min}(T_{SH}) = MSE(T_{ratio}) - \left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}^2(C_x - \rho_{yx}C_y)^2 \tag{17}$$

2.5 Singh and Deo (2003) estimator

This method uses power transformation in survey sampling. In this method, the data after imputation takes the following form:

$$y_{.i} = \begin{cases} y_i,\ if\ i \in R \\ \bar{y}_r\left[n\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta} - r\right]\frac{x_i}{\sum_{i\in\bar{R}} x_i},\ if\ i \in \overline{R} \end{cases} \tag{18}$$

The estimator of population mean is given by

$$T_{SD} = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta} \tag{19}$$

where β is an appropriate constant with optimum value $\beta^* = \rho_{yx}\frac{C_y}{C_x}$. The bias of T_{SD} is given by

$$Bias(T_{SD}) = \left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}\left(\frac{\beta(\beta-1)}{2}C_x^2 - \alpha\rho_{yx}C_yC_x\right) \tag{20}$$

Using β^*, we get the minimum MSE of T_{SD} as

$$MSE_{\min}(T_{SD}) = MSE(T_{ratio}) - \left(\frac{1}{r} - \frac{1}{n}\right)S_x^2\left(\frac{S_x^2}{S_{yx}} - \frac{\bar{Y}}{\bar{X}}\right)^2 \tag{21}$$

2.6 Singh (2009) estimator

This method is an alternative method of imputation to estimate population mean $\bar{Y}$ in the presence of non-response. The data take following form after imputation:

$$y_{.i} = \begin{cases} y_i, \; if\, i \in R \\ \bar{y}_r \left[\dfrac{(n-r)\bar{x}_n + \alpha r(\bar{x}_n - \bar{x}_r)}{\alpha \bar{x}_r + (1-\alpha)\bar{x}_n} \right] \dfrac{x_i}{\sum\limits_{i \in \bar{R}} x_i}, \; if\, i \in \overline{R} \end{cases} \tag{22}$$

Using this data, we obtain the estimator of population mean as following:

$$T_{Singh} = \frac{\bar{y}_r \bar{x}_n}{\gamma \bar{x}_r + (1-\gamma)\bar{x}_n} \tag{23}$$

where γ is an appropriate constant with optimum value $\gamma^* = \rho_{yx}\frac{C_y}{C_x}$. The bias of T_{Singh} is given by

$$Bias\left(T_{Singh}\right) = \bar{Y}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{yx}C_yC_x + \alpha^2\left(\frac{1}{r}-\frac{1}{n}\right)C_x^2 + (1-\alpha)^2\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2 - \alpha\left\{\left(\frac{1}{r}-\frac{1}{n}\right)\rho_{yx}C_yC_x + \left(\frac{1}{n}-\frac{1}{N}\right)C_x^2\right\} \right.$$
$$\left. +2\alpha(\alpha-1)\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2 - (1-\alpha)\left(\frac{1}{n}-\frac{1}{N}\right)\left(\rho_{yx}C_yC_x + C_x^2\right)\right] \tag{24}$$

Using γ^*, we get the minimum MSE of T_{Singh} as

$$MSE_{\min}(T_{Singh}) = MSE(T_{ratio}) - \left(\frac{1}{r}-\frac{1}{n}\right)S_x^2\left(\frac{S_{yx}}{S_x^2} - \frac{\bar{Y}}{\bar{X}}\right)^2 \tag{25}$$

2.7 Gira (2015) estimator

In this method, a ratio type imputation method was proposed such as

$$y_{.i} = \begin{cases} y_i & if\, i \in R \\ \bar{y}_r \left[n\left(\dfrac{\delta - \bar{x}_r}{\delta - \bar{x}_n}\right) - r \right] \dfrac{x_i}{\sum\limits_{i \in \bar{R}} x_i} & if\, i \in \bar{R} \end{cases} \tag{26}$$

where δ is an appropriate chosen constant in such a way that the MSE of the resultant estimator is minimum. Note that if $\delta = 0$ then $T_{Gira} = T_{ratio}$. The resultant estimator is obtained as

$$T_{Gira} = \bar{y}_r\left(\frac{\delta - \bar{x}_r}{\delta - \bar{x}_n}\right) \tag{27}$$

The bias of the above estimator is given by

$$Bias\left(T_{Gira}\right) = -\frac{\bar{X}\bar{Y}}{\delta - \bar{X}}\left(\frac{1}{r}-\frac{1}{n}\right)\rho_{yx}C_yC_x \tag{28}$$

Using the optimum value of $\delta = \bar{X}\left(\frac{C_x}{\rho_{yx}C_y} - 1\right)$, we get the minimum MSE of T_{Gira} as following:

$$MSE\left(T_{Gira}\right) = V\left(T_{mean}\right) - \left(\frac{1}{r}-\frac{1}{n}\right)\bar{Y}^2{\rho_{yx}}^2C_y^2 \tag{29}$$

3 The proposed method of imputation

A new method of imputation has been proposed to estimate the population mean. Using this proposed method, the data takes following form:

$$y_{.i} = \begin{cases} y_i, & if\, i \in R \\ \bar{y}_r \left[\dfrac{\{m(n+r)-r\}\bar{x}_r + \{(1-m)n - mr\}\bar{x}_n}{m\bar{x}_n + (1-m)\bar{x}_r} \right] \dfrac{x_i}{\sum\limits_{i\in\bar{R}} x_i}, & if\, i \in \overline{R} \end{cases} \tag{30}$$

The point estimator of population mean under this method comes out as:

$$T_p = \bar{y}_r \left[\frac{m\bar{x}_r + (1-m)\bar{x}_n}{m\bar{x}_n + (1-m)\bar{x}_r} \right] \tag{31}$$

where $m\,(\neq 0)$ is some real number.

4 Properties of the proposed estimator

We define $\bar{y}_r = \bar{Y}(1+e_0)$, $\bar{x}_r = \bar{X}(1+e_1)$, $\bar{x}_n = \bar{X}(1+e_2)$. We get, $E(e_i) = 0;\ \forall i = 0,1,2$ and

$$E\left(e_o^2\right) = \left(\frac{1}{r} - \frac{1}{N}\right) C_y^2,\ E\left(e_1^2\right) = \left(\frac{1}{r} - \frac{1}{N}\right) C_x^2,\ E\left(e_2^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right) C_x^2$$

where $C_x^2 = \frac{S_x^2}{\bar{X}}$, $C_y^2 = \frac{S_y^2}{\bar{Y}}$ are the population coefficients of variation of auxiliary and study variables respectively, $\rho_{yx} = \frac{S_{xy}}{S_x S_y}$ is the coefficient of correlation between auxiliary and study variable and S_x^2, S_y^2 have their usual meaning. Using above transformations in proposed method, we get

$$T_p = \bar{Y}(1+e_0) \left[\frac{m\bar{X}(1+e_1) + (1-m)\bar{X}(1+e_2)}{m\bar{X}(1+e_2) + (1-m)\bar{X}(1+e_1)} \right] \tag{32}$$

After some algebraic calculations, we get

$$Bias(T_p) = \bar{Y}\left(\frac{1}{r} - \frac{1}{n}\right)\left[\left(2m^2 + 1 - 3m\right) C_x^2 + (2m-1)\rho_{yx} C_y C_x\right] \tag{33}$$

and

$$MSE(T_p) = \bar{Y}^2 \left[\left(\frac{1}{r} - \frac{1}{N}\right) C_y^2 + (1-2m)^2 \left(\frac{1}{r} - \frac{1}{n}\right) C_x^2 - 2(1-2m)\left(\frac{1}{r} - \frac{1}{n}\right)\rho_{yx} C_y C_x \right] \tag{34}$$

Minimizing $MSE(T_p)$ with respect to m, we get $m_{opt} = \frac{1}{2}\left(1 - \rho_{yx}\frac{C_y}{C_x}\right)$. Utilizing this optimum value, we get minimum of $MSE(T_p)$ as following:

$$Min.MSE(T_p) = \bar{Y}^2 \left[\left(\frac{1}{r} - \frac{1}{N}\right) C_y^2 - \left(\frac{1}{r} - \frac{1}{n}\right) {\rho_{yx}}^2 {C_y}^2 \right] \tag{35}$$

5 Practicability

The choice of m is the main problem in using the suggested imputation method. We must note that the optimum value of m is received in terms of the very familiar parameter $K = \rho_{yx}(C_y/C_x)$. The value of K is quite stable in the repeated surveys as shown by [13]. Thus, if the value of K is given then the suggested method can be easily implemented in actual surveys. Many times, the value of K is unknown to us. In such situations, we recommend two estimators of m, given by

$$\hat{m}_1 = \frac{1}{2}\left(1 - \frac{\bar{x}_n s_{xy}^*}{\bar{y}_r s_{x(r)}^2}\right) \tag{36}$$

$$\hat{m}_2 = \frac{1}{2}\left(1 - \frac{\bar{x}_n s^*_{xy}}{\bar{y}_r s^2_{x(n)}}\right) \tag{37}$$

where $s^*_{xy} = \frac{1}{r-1}\sum_{i\in R}(y_i - \bar{y}_r)(x_i - \bar{x}_r)$, $s^2_{x(r)} = \frac{1}{r-1}\sum_{i\in s}(x_i - \bar{x}_r)^2$ and $s^2_{x(n)} = \frac{1}{n-1}\sum_{i\in s}(x_i - \bar{x}_r)^2$ Thus, to get a practicable estimator of population mean, the unknown parameter m in Equation (31) can be replaced with its estimators $\hat{m}_1$ and $\hat{m}_2$ from Equations (36) and (37). The asymptotic MSE of the resultant estimator of population mean remains the same after replacing m with $\hat{m}_1$ or $\hat{m}_2$ ([14]). Hence, the proposed method of imputation remains better than the ratio and mean methods of imputation.

6 Remark

It is also possible to propose the following imputation methods and their resultant point estimators of population mean in two different strategies

6.1 Under this method, the data becomes

$$y_{.i} = \begin{cases} y_i, & if\, i \in R \\ \bar{y}_r \left[\frac{\{m(n+r)-r\}\bar{x}_r + \{(1-m)n - mr\}\bar{X}}{m\bar{X} + (1-m)\bar{x}_r}\right] \frac{x_i}{\sum_{i\in\bar{R}} x_i}, & if\, i \in \overline{R} \end{cases} \tag{38}$$

and the resultant point estimator is obtained as

$$T_1 = \bar{y}_r \left[\frac{m\bar{x}_r + (1-m)\bar{X}}{m\bar{X} + (1-m)\bar{x}_r}\right] \tag{39}$$

6.2 Under this method, the data becomes

$$y_{.i} = \begin{cases} y_i, & if\, i \in R \\ \bar{y}_r \left[\frac{\{m(n+r)-r\}\bar{x}_n + \{(1-m)n - mr\}\bar{X}}{m\bar{X} + (1-m)\bar{x}_n}\right] \frac{x_i}{\sum_{i\in\bar{R}} x_i}, & if\, i \in \overline{R} \end{cases} \tag{40}$$

and the resultant point estimator is obtained as

$$T_2 = \bar{y}_r \left[\frac{m\bar{x}_n + (1-m)\bar{X}}{m\bar{X} + (1-m)\bar{x}_n}\right] \tag{41}$$

7 Numerical demonstration

Using four different data sets given in Table 1, we demonstrate the performance of the proposed imputation method over some contemporary methods of imputation for different values of n and r. The percent relative efficiencies (PREs) of the proposed estimator with respect to the mean method of imputation (T_{mean}), ratio method of imputation (T_{ratio}), regression method of imputation (T_{reg}), compromised method of imputation (T_{SH}), Singh and Deo (2003) estimator (T_{SD}), Singh (2009) estimator (T_{Singh}) and Gira (2015) estimator (T_{Gira}), respectively are calculated as followings and results are presented in Tables 2 and 3.

$$PRE_1 = \frac{V(T_{mean})}{MSE(T_p)} \times 100 \tag{42}$$

$$PRE_2 = \frac{MSE(T_{ratio})}{MSE(T_p)} \times 100 \tag{43}$$

$$PRE_3 = \frac{MSE\left(T_{reg}\right)}{MSE\left(T_p\right)} \times 100 = \frac{MSE\left(T_{SH}\right)}{MSE\left(T_p\right)} \times 100 = \frac{MSE\left(T_{SD}\right)}{MSE\left(T_p\right)} \times 100 = \frac{MSE\left(T_{Singh}\right)}{MSE\left(T_p\right)} \times 100 = \frac{MSE\left(T_{Gira}\right)}{MSE\left(T_p\right)} \times 100 \quad (44)$$

Table 1 Population parameters of four different populations.

Parameters	Population I (Kadilar and Cingi (2008))	Population II (Singh (2009))	Population III (Diana and Perri (2010))	Population IV (Lohr (1999))
N	19	3055	8011	3059
n	10	611	400	61, 122, 183, 244, 305, 367, 428, 489, 550, 611
r	8	520	360	(6, 12, 18, 24,30), (12, 24, 36, 48, 60), (18, 36, 54, 72, 90), (24, 48, 72, 96, 120), (30, 60, 90, 120, 150), (36, 72, 108, 144, 180), (42, 84, 126, 168, 210), (48, 96, 144, 192, 240), (55, 110, 165, 220, 275), (61,122, 183, 244, 305)
$\bar{Y}$	575.00	308582.4	28229.43	308582.4
$\bar{X}$	13537.68	56.5	1.69	56.5
S_y	858.36	425312.8	22216.56	425312.8
S_x	12945.38	72.3	0.78	72.26842
ρ_{yx}	0.88	0.677	0.46	0.6774282

Table 2 *PREs* of the proposed estimators under three different population data sets.

Population Source	Population I (Kadilar and Cingi (2008))	Population II (Singh (2009))	Population III (Diana and Perri (2010))
PRE_1	136.5225	108.9638	102.2658
PRE_2	102.7036	101.2364	100.1712
PRE_3	100	100	100

Table 3 *PREs* of the considered estimators under simulation study using population 4.

n	r	PRE_1	PRE_2	PRE_3	n	r	PRE_1	PRE_2	PRE_3
61	6	158.868	136.267	100.697	367	36	168.362	113.729	101.332
	12	152.127	118.277	102.782		72	158.831	110.398	100.379
	18	143.245	112.977	100.404		108	150.141	107.857	99.982
	24	136.457	109.223	101.807		144	141.319	106.263	100.182
	30	128.961	106.592	100.164		180	132.369	104.913	100.313
122	12	164.881	122.442	100.983	428	42	169.179	113.400	99.280
	24	154.324	113.762	101.786		84	159.939	109.570	99.572
	36	145.938	109.639	101.865		126	149.308	107.883	99.943
	48	137.639	107.569	100.418		168	140.857	106.513	100.256
	60	130.77	105.466	99.918		210	133.439	105.148	100.305
183	18	162.545	118.281	99.195	489	48	171.955	112.779	100.322
	36	155.891	111.807	102.866		96	160.604	109.452	99.743
	54	146.941	108.938	100.269		144	151.674	107.643	102.077
	72	139.296	106.682	99.806		192	141.499	106.359	100.184
	90	130.942	105.178	100.070		240	133.828	104.851	100.326
244	24	165.485	116.247	99.147	550	55	168.374	113.040	100.722
	48	157.358	110.964	101.963		110	160.004	109.492	99.829
	72	148.764	108.199	100.747		165	150.339	107.421	100.162
	96	139.416	106.495	99.973		220	142.517	106.048	100.275
	120	131.841	104.836	100.255		275	132.565	105.301	100.273
305	30	167.238	114.507	100.195	611	61	170.354	112.312	99.049
	60	158.313	110.023	100.480		122	161.735	109.496	99.762
	90	147.967	108.091	99.771		183	150.626	108.091	100.067
	120	139.264	106.627	100.020		244	141.599	106.540	100.126
	150	132.635	104.773	100.263		305	133.838	105.085	100.397

8 Simulation study

Using Population 4 ([15]) given in Table 1, simulation study is performed for the sample population from 2% to 20% with response rate from 10% to 50%. From this population, the study variable Y_i = number of acres devoted to farms during 1992 (acres92) and the auxiliary variable X_i = number of large farms during 1992 (largef92) have been taken. The data set given in file agpop.dat has been used in these illustrations after dropping the data values marked as -99. The descriptive parameter of the study variable acres92 and auxiliary variable largef92 after cleaning the data, which is used here, is given in Table 1. Throughout this study, calculations are used based on simulation of 10,000 repeated samples without replacement.

9 Discussions

From Tables 2 and 3, following interpretations can be drawn:
(1) From Table 2, we can easily observe that for 5% to 50% sample population, response rate between 80% and 90% having different correlation coefficient, the proposed estimator remains better than mean method and ratio method of imputation.

(2) Based on the results of Table 3, it may be observed that for 2% - 20% of sample fraction with different response rate in each case, the PREs of the proposed estimator with respect to the estimators in mean method and ratio method of imputation remains more than 100% and it gives approximately equal result with respect to regression, Singh and Horn (2000), Singh and Deo (2003), Singh (2009) and Gira (2015) method of imputation.

10 Conclusions

In this paper, a different method of imputation and the resultant point estimator of population mean is obtained which gives better result than the traditional estimators (mean method of imputation and ratio method of imputation) in case of missing data. The percent relative efficiency implies that the proposed estimator is equivalent to the other discussed estimators.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

Conflicts of Interests
The authors declare that they have no conflicts of interests

References

[1] D. B. Rubin, Inference and missing data. *Biometrika*, **63**, 581-592 (1976).
[2] D. F. Heitjan and S. Basu, Distinguishing "missing at random" and "missing completely at random". *The American Statistician*, **50**, 207-213 (1996).
[3] A. A. Gira, Estimation of population mean with a new imputation methods. *Applied Mathematical Sciences*, **9**, 1663-1672 (2015).
[4] C. Kadilar and H. Cingi, Estimators for the population mean in the case of missing data. *Communications in Statistics — Theory and Methods*, **37**, 2226-2236 (2008).
[5] G. Diana, and P. Francesco Perri, Improved estimators of the population mean for missing data. *Communications in Statistics—Theory and Methods*, **39**, 3245-3251 (2010).
[6] G. N. Singh, K. Priyanka, J. M. Kim and S. Singh, Estimation of population mean using imputation techniques in sample surveys. *Journal of the Korean Statistical Society*, **39**, 67-74 (2010).
[7] S. Bhushan and A. P. Pandey, Optimal imputation of missing data for estimation of population mean. *Journal of Statistics and Management Systems*, **19**, 755-769 (2016).
[8] S. Prasad, Ratio exponential type estimators with imputation for missing data in sample surveys. *Model Assisted Statistics and Applications*, **12**, 95-106 (2017).
[9] S. Singh, A new method of imputation in survey sampling. *Statistics*, **43**, 499-511 (2009).
[10] S. Singh and B. Deo, Imputation by power transformation. *Statistical Papers*, **44**, 555-579 (2003).
[11] S. Singh and S. Horn, Compromised imputation in survey sampling. *Metrika*, **51**, 267-276 (2000).
[12] H. Lee,, E. Rancourt and C. E. Saerndal, Experiments with variance estimation from survey data with imputed values. *Journal of official statistics*, **10**, 231-243 (1994).
[13] V. N. Reddy, A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya C*, **40**, 29-37 (1978).
[14] S. Sampath, On the optimal choice of unknowns in ratio type estimators. *J. Indian Soc. Agric. Stat*, **41**, 166-172 (1989).
[15] S.L. Lohr, *Sampling: Design and Analysis*, Duxbury Press, Pacific Grove CA, 221-249, (1999).