Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/090202

Estimation of Mean with Imputation of Missing data using Exponential-Type Estimators

Ajeet Kumar Singh[1,*], *Priyanka Singh*[2] *and V. K. Singh*[3]

[1]Department of Statistics, University of Rajasthan, Jaipur, India
[2]SRMIST Kattankulathur, Chennai, India
[3]Department of Statistics, Banaras Hindu University, India

Received: 15 Jul. 2021, Revised: 2 Sep. 2021, Accepted: 24 Nov. 2021
Published online: 1 May 2022

Abstract: There are several methods of handling missing data in sample surveys, which is a typical problem of non-response. Imputation method is one of the methods to deal with non-response. In this paper suggests a one-parameter family of estimators, popularly known as Exponential-Type Estimator (ETE), use as a tool of imputation for dealing with non-responding units and discuss its properties. The proposed strategy has been observed to be more precise than other method of imputation such as mean, ratio and compromised method of imputation under optimality conditions. To support the discussed results, a numerical illustration has been performed on data set.

Keywords: Imputation methods, one-parameter family of estimators, optimum estimator, bias and mean square error.

1 Introduction

Imputation refers to the process of assigning one or more values to an item when there is no reported value for that item. Many forms of imputation are available, mean method, hot-deck imputation, ratio imputation, regression imputation are all single imputation in the sense that a single value is imputed for every missing value to produce a complete data set. To deal with missing values effectively Kalton *et al* (1981) and Sande (1979) 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 variate if it is available. Lee *et al* (1994) used the information on an auxiliary variable for the purpose of imputation. Based on auxiliary variable, recently Singh and Horn (2000), Ahmed *et al* (2006), Shakti Prasad (2017) and Singh *et al* (2014a, 2014b, 2015, 2016a, 2016b) suggested some method of imputation.

2 The Problem and Notations

Let a simple random sample S of size n without replacement be drawn from a finite population $U = (Y_1, Y_2, \ldots, Y_N)$ of size N and with study characteristic Y. Let $(\bar{Y}, \bar{X})$ be the population mean of the study variable Y and auxiliary variable X respectively. It is presumed that the sample consists of r responding units (r < n) belonging to a set A and (n-r) non -responding units belonging to a set A^C. Further let for every unit $i \in A$, the value y_i is observed and for the unit $i \in A^C$, the value y_i is missing for which suitable imputed value is to be derived. For this purpose, the ith value of the auxiliary variable is used as a source of imputation for missing data when $i \in A^C$.

In what follows, we shall use the following notations:
Z : Stands for either variable Y or variable X.
$\bar{z}_n$:Sample mean based on n observations for variable Z.
$\bar{z}_r$:Sample mean of the responding units based on r observations for the variable Z.

[*]Corresponding author e-mail: ajeetvns.singh@gmail.com

S_Z^2 :Population mean squares for the variable Z.

C_Z: Coefficient of variation (CV) for the variable Z; $C_Z = \frac{S_Z}{\overline{Z}}$.

ρ is the coefficient of correlation between the variable Y and X in the population .

$y_{.i}$: Imputed value for the ith value of $y_i (i = 1,2...n)$

$\theta_{n,N}, \theta_{r,N}, \theta_{r,n}$:Finite population corrections (fpc); $\left(\frac{1}{n}-\frac{1}{N}\right), \left(\frac{1}{r}-\frac{1}{N}\right), \left(\frac{1}{r}-\frac{1}{n}\right)$respectively.

3 Some Imputation Strategies

Before suggesting the proposed imputation strategy, we shall mention here some existing imputation strategies for readiness of the material which has a direct relevance with the present work. We shall denote by (D, T) denote a sampling strategy where D stands for simple random sampling without replacement sampling scheme and T for an estimator for population mean $\overline{Y}$. Followings are the some imputation methods and corresponding sampling strategies:

3.1 $(D, \bar{y}_r)$:***Mean method***

Here

$$y_{.i} = \begin{cases} y_i & \text{if } \mathrm{i} \in \mathrm{A} \\ \bar{\mathrm{y}}_{\mathrm{r}} & \text{if } \mathrm{i} \in \mathrm{A}^{\mathrm{C}} \end{cases} \tag{1}$$

The corresponding point estimator and its bias, B(.) and mean square error(MSE),M(.) are derived as

$$\bar{\mathrm{y}}_{\mathrm{s}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i} \in \mathrm{s}} \mathrm{y}_{.\mathrm{i}} = \bar{\mathrm{y}}_{\mathrm{r}} \tag{2}$$

$$B(\bar{y}_r) = 0 \tag{3}$$

$$M(\bar{y}_r) = \theta_{r,N} \overline{Y}^2 C_Y^2 \tag{4}$$

3.2 $(D, \bar{y}_{RAT})$***: Ratio method***

$$y_{.i} = \begin{cases} y_i & \text{if } \mathrm{i} \in \mathrm{A} \\ \hat{\mathrm{b}}\mathrm{x}_{\mathrm{i}} & \text{if } \mathrm{i} \in \mathrm{A}^{\mathrm{C}} \end{cases} \tag{5}$$

where $\hat{b} = \frac{\sum_{i \in A} y_i}{\sum_{i \in A} x_i}$

Then the point estimator, its bias and MSE are given by:

$$\bar{\mathrm{y}}_{\mathrm{RAT}} = \bar{\mathrm{y}}_{\mathrm{r}} \frac{\bar{\mathrm{x}}_{\mathrm{n}}}{\bar{\mathrm{x}}_{\mathrm{r}}} \tag{6}$$

$$B(\bar{y}_{RAT}) = \theta_{r,n} \overline{Y} \left[C_X^2 - \rho_{XY} C_X C_Y\right] \tag{7}$$

$$M(\bar{y}_{RAT}) = \theta_{r,N} \overline{Y}^2 C_Y^2 + \theta_{r,n} \overline{Y}^2 \left[C_X^2 - 2\rho_{XY} C_X C_Y\right] \tag{8}$$

3.3 $(D, \bar{y}_{COMP})$***: Compromised method*** (Singh and Horn, 2000)

$$y_{.i} = \begin{cases} \alpha \frac{n}{r} y_i + (1\text{-}\alpha)\hat{\mathrm{b}}\mathrm{x}_{\mathrm{i}} & \text{if } \mathrm{i} \in \mathrm{A} \\ (1\text{-}\alpha)\,\hat{\mathrm{b}}\mathrm{x}_{\mathrm{i}} & \text{if } \mathrm{i} \in \mathrm{A}^{\mathrm{C}} \end{cases} \tag{9}$$

The point estimator is

$$\bar{\mathrm{y}}_{\mathrm{COMP}} = \alpha\bar{\mathrm{y}}_{\mathrm{r}} + (1-\alpha)\bar{\mathrm{y}}_{\mathrm{r}} \frac{\bar{\mathrm{x}}_{\mathrm{n}}}{\bar{\mathrm{x}}_{\mathrm{r}}} \quad ;\alpha \text{ being a suitable constant} \tag{10}$$

$$B(\bar{y}_{COMP}) = (1-\alpha)\theta_{r,n}\bar{Y}\left[C_X^2 - \rho_{XY}C_X C_Y\right] \tag{11}$$

$$M(\bar{y}_{COMP}) = \theta_{r,N}\bar{Y}^2 C_Y^2 + \theta_{r,n}\bar{Y}^2\left[(1-\alpha)^2 C_X^2 - 2(1-\alpha)\rho_{XY}C_X C_Y\right] \tag{12}$$

It can be seen that the estimator has minimum MSE for $\alpha = 1 - \rho\dfrac{C_Y}{C_X}$ for which

$$M(\bar{y}_{COMP})_{\min} = \bar{Y}^2\left[(\theta_{r,N} - \theta_{r,n}\rho^2)C_Y^2\right] \tag{13}$$

4 Proposed Imputation Strategies $(D, y_{T_i}), i = 1,2,3$

Motivated with Singh *et al* (2014)and Bahl and Tuteja (1991), we here proposed the following exponential-type estimators

4.1

$$(y_{.i}) = \begin{cases} y_i & if \quad i \in A \\ \dfrac{\bar{y}_r}{(n-r)}\left[nT_1 - r\right] & if \quad i \in A^C \end{cases} \tag{14}$$

4.2

$$(y_{.i}) = \begin{cases} y_i & if \quad i \in A \\ \dfrac{\bar{y}_r}{(n-r)}\left[nT_2 - r\right] & if \quad i \in A^C \end{cases} \tag{15}$$

4.3

$$(y.i) = \begin{cases} y_i & if \quad i \in A \\ \dfrac{\bar{y}_r}{(n-r)}\left[nT_3 - r\right] & if \quad i \in A^C \end{cases} \tag{16}$$

where

$$T_1 = \exp\left[\alpha\left(\frac{\bar{X} - \bar{x}_n}{\bar{X} + \bar{x}_n}\right)\right] \tag{17}$$

$$T_2 = \exp\left[\alpha\left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right)\right] \tag{18}$$

$$T_3 = \exp\left[\alpha\left(\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right)\right] \tag{19}$$

where α is a suitably chosen constant to be determined under certain conditions.

Under equations (17), (18) and (19) the point estimator of $\bar{Y}$ are

$$\left(y_{T_1}\right)=\bar{\mathrm{y}}_{\mathrm{r}}\; T_1 \tag{20}$$

$$\left(y_{T_2}\right)=\bar{y}_r T_2 \tag{21}$$

$$\left(y_{T_3}\right)=\bar{\mathrm{y}}_{\mathrm{r}}\; T_3 \tag{22}$$

Remark 1: It is clear that T_i (i = 1, 2, 3) defines classes of estimators if the parameter α takes different values. α is a suitably chosen constant whose value can be determined under certain conditions. Some important value of α are 0, 1 and -1 for which (i = 1, 2, 3) reduces to 1, a ratio-type estimator and a product-type estimator respectively. Accordingly, the proposed imputation schemes reduce to mean imputation, ratio method type imputation and product method type imputation respectively.

5 Properties of proposed Imputation Strategies

In relation to bias, MSE, optimum value of the parameter α and corresponding minimum MSE, we have the following theorems:

5.1*Theorem1:* The bias and MSE of the proposed strategy (D, y_{T_1}) to the terms of order $O(n^{-1})$ are given by

$$B\left(y_{T_1}\right)=\theta_{n,N}\bar{Y}\left(\frac{\alpha}{4}C_X^2+\frac{\alpha^2}{8}C_X^2-\frac{\alpha}{2}\rho C_Y C_X\right) \tag{23}$$

$$M(y_{T_1})=\bar{Y}^2\left[\theta_{r,N}C_Y^2+\theta_{n,N}\frac{\alpha^2}{4}C_X^2-\alpha\theta_{n,N}\rho C_Y C_X\right] \tag{24}$$

The minimum MSE of $\left(y_{T_1}\right)$ occurs when $\alpha = 2\rho\frac{\mathrm{C_Y}}{\mathrm{C_X}}$ for which MSE reduces to

$$M(y_{T_1})_{Min}=\bar{Y}^2\left[(\theta_{r,N}-\theta_{n,N}\rho^2)C_Y^2\right] \tag{25}$$

The proof of the theorem is given in the Appendix.

5.2Theorem2: The bias and MSE of the proposed strategy (D, y_{T_2}) to the terms of order $O(n^{-1})$ are given by

$$B\left(y_{T_2}\right)=\bar{Y}\theta_{r,n}\left(\frac{\alpha}{4}C_X^2+\frac{\alpha^2}{8}C_X^2-\frac{\alpha}{2}\rho C_Y C_X\right) \tag{26}$$

$$M(y_{T_2})=\bar{Y}^2\left[\theta_{r,N}C_Y^2+\theta_{r,n}\frac{\alpha^2}{4}C_X^2-\alpha\theta_{r,n}\rho C_Y C_X\right] \tag{27}$$

Minimum MSE occurs at $\alpha = 2\rho\frac{\mathrm{C_Y}}{\mathrm{C_X}}$ and the corresponding expression for MSE will be

$$M(y_{T_2})_{Min}=\bar{Y}^2\left[(\theta_{r,N}-\theta_{r,n}\rho^2)C_Y^2\right] \tag{28}$$

which is same as that of the strategy $(D, \bar{y}_{COMP})$ under optimality condition.
The proof of the theorem is given in the Appendix.

5.3*Theorem3:* The bias and MSE of the proposed strategy (D, y_{T_3}) to the terms of order $O(n^{-1})$ are given by

$$B\left(y_{T_3}\right)=\theta_{r,N}\bar{Y}\left[\left(\frac{\alpha}{4}+\frac{\alpha^2}{8}\right)C_X^2-\frac{\alpha}{2}\rho C_Y C_X\right] \tag{29}$$

$$M(y_{T_3}) = \overline{Y}^2 \theta_{r,N}\left[C_Y^2 + \frac{\alpha^2}{4}C_X^2 - \alpha\rho C_Y C_X\right] \tag{30}$$

Minimum MSE occurs at $\alpha = 2\rho\frac{C_Y}{C_X}$ for which MSE expression reduces to

$$M(y_{T_3})_{Min} = \overline{Y}^2\left[\theta_{r,N}C_Y^2(1-\rho^2)\right] \tag{31}$$

The proof of the theorem is given in the Appendix.

Remark2: It is interesting that the MSEs of all the strategies are optimum for the same value of the parameter α, but the optimum estimators in each strategy are having different values of MSE. It is, therefore, a matter of interest to observe the best strategy among the proposed strategies as well as to make a comparison between the strategies for a random choice of α. Also, it is important to observe that how the suggested strategies out performs the strategies $[D, \overline{y}_r]$, $[D, \overline{y}_{RAT}]$ and $[D, \overline{y}_{COMP}]$ in general.

6 Comparison of Different Strategies

It is now desirable to compare the three suggested strategies for their performances .We have

$$\textbf{(i)}D_1 = M(y_{T_2})_{Min} - M(y_{T_1})_{Min} = \left(\theta_{n,N} - \theta_{r,n}\right)\rho^2 C_Y^2 \tag{32}$$

Thus y_{T_1} would be preferable over y_{T_2} if

$$\text{if } r > \frac{n}{2-f} \;;\quad \text{where f} = \frac{n}{N} \quad (0 < f < 1) \tag{33}$$

Thus, theoretically

that is, when the non-response in the sample is observed to be less than fifty percent. In any kind of survey, it generally holds, and, therefore, it could be expected that the strategy y_{T_1} would be fairly better than strategy y_{T_2} under the optimality conditions, in most of the survey situations. From the expression (33), it can, therefore, be concluded that the strategy y_{T_1} **would be better than** y_{T_2} **under the optimality condition if as the sample size tends to be larger, the number of non-responding units in the sample is smaller and smaller.**

$$\textbf{(ii)}D_2 = M(y_{T_1})_{Min} - M(y_{T_3})_{Min} = \theta_{r,n}\rho^2 C_Y^2 \tag{34}$$

which is always positive. Therefore strategy (D, y_{T_3}) is always better than strategy (D, y_{T_1}).

$$\textbf{(iii)}D_3 = M(y_{T_2})_{Min} - M(y_{T_3})_{Min} = \theta_{n,N}\rho^2 C_Y^2 \tag{35}$$

which is always positive. Therefore strategy (D, y_{T_3}).is always better than strategy (D, y_{T_2}).

Combining the results derived above, one can say that if the sample does not contain more than fifty percent non – respondents, then the following results holds.

$$M(y_{T_3}) \le M(y_{T_1}) \le M(y_{T_2})$$

6.2 Since for the proposed strategies, the value of the parameter α may assume positive as well as negative values, the comparison of the three strategies for an arbitrary choice of α will include a number of conditions for the choice of α so that one strategy would be preferable over the others. Hence, it would be of no use to compare the strategies theoretically, rather these might be compared on the basis of some empirical data. Due to this reason, for the comparison purpose, it is always preferable to select the choice of α as α_0 everywhere.

7 Empirical Study

In order to illustrate the results observed theoretically, we have considered two data sets, which have been described below:

7.1 Population I: We consider the data given in Mukhopadhyay (2000). The population consists of 20 jute mills. The data show the numbers of labourers X (in thousands) and quantity of raw materials required Y (in lakhs of bales). Here we take n = 7 and r = 5. For the given population, we have the following values: $\overline{Y}$ = 41.5, $\overline{X}$ = 441.95, S_X^2 = 10215.21, S_Y^2 = 95.7368, ρ = 0.6521. The values of bias and minimum MSE of the suggested strategies are depicted in the following table. However for comparison purpose, we have also shown the corresponding bias and MSE of the strategies $(D, \overline{y}_r), (D, \overline{y}_{RAT})$, and $(D, \overline{y}_{COMP})$ in the table and the percent relative efficiency of the strategies with respect to $(D, \overline{y}_r)$ The bias and MSE of the strategy $(D, \overline{y}_{COMP})$ has been obtained under the corresponding optimality conditions.

7.2 Population II: This population has been borrowed from the work of Shukla et al (2011b). They generated an artificial data set of size N = 200 which consists of pairs of values (yi, xi) (i = 1, 2,...,200) of the two variables Y and X; Y being the study variable and X being the auxiliary variable. The values of required population parameters were obtained as follows:

N = 200, $\overline{Y}$ = 42.485, $\overline{X}$ = 18.515, S_Y = 14.1088, S_X = 6.9669, ρ = 0.8652.

For the purpose of calculation of bias and MSE of the estimators we take n = 20 and r = 15. Similar to Table 1, the following table depicts the results related to various strategies for this population:

Table 1. Bias, MSE and Percent Relative Efficiency of the strategies

Strategy	Bias	MSE	PRE
$(D, \overline{y}_r)$	0.00	14.361	100.00
$(D, \overline{y}_{RAT})$	0.04	12.589	114.08
$(D, \overline{y}_{COMP})$	0.14	12.034	119.34
(D, y_{T_1})	0.02	10.583	135.70
(D, y_{T_2}).	0.01	12.034	119.34
(D, y_{T_3})	-0.11	8.256	173.95

Table 2. Bias, MSE and Percent Relative Efficiency of the strategies

Strategy	Bias	MSE	PRE
$(D, \overline{y}_r)$	0.00	12.276	100.00
$(D, \overline{y}_{RAT})$	0.02	9.846	124.68

$(D, \bar{y}_{COMP})$	0.02	9.608	127.77
(D, y_{T_1})	0.02	5.570	220.39
(D, y_{T_2}).	0.01	9.792	125.36
(D, y_{T_3})	0.02	3.087	397.73

Remark 3: The trends of the results obtained in Table 2, are almost similar to that observed for Population I in Table 1.

8 Simulation Study

8.1

It is obvious that whenever the efficiency of one strategy is calculated with respect to another strategy, such a comparison requires the known population values. Therefore, such an efficiency comparison has very limited practicability. In order to avoid this difficulty, sample survey researchers now depend upon simulation studies which are based on actually drawn samples and thereby, sample values which are drawn from the given population. In the present age of high speed computers, a large number of samples of same size is not a difficult thing. We have, therefore, compared the performance of proposed strategies with each other and with other strategies on the basis of a simulation study as follows: The simulation procedure for the population I contains following steps:

Step 1: Select all the possible random samples of size 7 from the population of size 20, considered above, using SRSWOR. We get ${}^{20}C_7 = 77520$.

Step 2: Drop down 2 units randomly from each of the selected samples for which Y values are treated to be missing.

Step 3: Compute and impute the dropped units of *Y* with the help of proposed methods of imputation.

Step 4: Repeat the above steps 77,520 times, which provides multiple sample based estimates.

Step 5: Define the bias of the estimator $\hat{T}$ as

$$B\left(\hat{T}\right)=\frac{1}{77520}\sum_{i=1}^{77520}\left[\left(\hat{T}_i\right)-\bar{Y}\right]$$

Step 6: Define the M.S.E. of estimator T as

$$M[T]=\frac{1}{77520}\sum_{i=1}^{77520}\left[\left(\hat{T}_i\right)-\bar{Y}\right]^2$$

Table 3. Below depicts the bias MSE and PRE of other strategies with respect to the strategy $[D, \bar{y}_r]$ based on simulation study.

Table 3. Simulation based bias, MSE and PRE of different strategies.

Strategy	Bias	MSE	PRE
$[D, \bar{y}_r]$	0.0217	14.526	100.000
$[D, \bar{y}_{RAT}]$	0.0346	12.515	116.060
$[D, \bar{y}_{COMP}]$	0.0354	12.228	118.780
$[D, y_{T_1}]$	0.0223	10.416	139.451

$[\mathrm{D}, y_{T_2}]$.	0.0047	11.958	121.470
$[D, y_{T_3}]$	0.0397	8.181	177.540

8.2

Considering the population II, we have the following steps in the simulation study :

Step I: A random samples of size 20 from the population was selected, using SRSWOR.
Step II: Drop down 5 units randomly from each of the selected samples for which Y values are treated to be missing.
Step III: Compute and impute the dropped Y- values with the help of proposed methods of imputation and with the help of other methods of imputation under consideration.
Step IV: Repeat the above steps 30000 times, which provide multiple sample based estimates.
Step V: Define the bias of the estimator $\hat{T}$ as

$$B\left(\hat{T}\right) = \frac{1}{30000}\sum_{i=1}^{30000} \left[\left(\hat{T}_i\right) - \overline{Y}\right]$$

Step 6: Define the M.S.E. of estimator T as

$$M[T] = \frac{1}{30000}\sum_{i=1}^{30000} \left[\left(\hat{T}_i\right) - \overline{Y}\right]^2$$

Table 4. Simulation based bias, MSE and PRE of different strategies.

Strategy	Bias	MSE	PRE
$[D, \bar{y}_r]$	0.01	12.399	100.00
$[D, \bar{y}_{RAT}]$	0.02	9.756	127.10
$[D, \bar{y}_{COMP}]$	0.01	9.526	129.75
$[D, y_{T_1}]$	0.02	5.322	232.99
$[\mathrm{D}, y_{T_2}]$	0.03	9.640	128.62
$[D, y_{T_3}]$	0.04	2.850	435.13

Remark 5: Here, we have compared all the strategies given in Singh, G.N. *et al* (2016). We found that our Strategy $[D, y_{T_3}]$ is uniformly better than the other strategies and rest of our strategies is not much better than Singh, G.N. *et al* (2016) but some are approximately same.

9 Conclusions

From the table, it is evident that

(i) Strategy $[D, y_{T_3}]$ is uniformly better than the other strategies.

(ii) Strategy $[D, y_{T_1}]$ can be preferred over strategies $[D, \bar{y}_r]$, $[D, \bar{y}_{RAT}]$ and $[D, \bar{y}_{COMP}]$.

(iii) Strategy $[\mathrm{D}, y_{T_2}]$ is equally efficient to the compromised method of imputation suggested by Singh and Horn (2000). It can therefore, be concluded that proposed methods of imputation have better performance than the previously developed methods $[D, \bar{y}_r]$, $[D, \bar{y}_{RAT}]$ and $[D, \bar{y}_{COMP}]$.

Acknowledgement: The authors are thankful to reviewers as well as the editor of the journal.

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] Bahl, S. and Tuteja, R. K. (1991): Ratio and product type exponential estimator, Information and Optimization sciences,Vol,XII,I,159-163.
[3] Kalton, G., Kasprzyk, D. and Santos, R. (1981): Issues of Non-Response and Imputation in the Survey of Income and Programme Participation, in Current Topics in Survey Sampling (eds.D.Krewski,R.Platek and J. N. K. Rao), Academic Press, New York,455-480.
[4]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.
[5] Mukhopadhyay, P. (2000): Theory and methods of survey sampling. prentice Hall of India Pvt.Ltd., New Delhi.
[6] Rao, J. N. K. and Sitter, R. R. (1995): Variance estimation under two phase sampling with application to imputation for missing data. Biometrika,82, 453−460.
[7] Sande, I. G. (1979): A personal view of hot deck imputation procedures, Survey Methodology, 5, 238−246.
[8] Singh, S. and Horn, S. (2000): Compromised imputation in survey sampling, Metrika, 51,267−276.
[9] Singh, A. K., Singh, P. and Singh, V. K. (2014a): Exponential-Type Compromised Imputation in Survey Sampling, Journal of the Statistics Applications and Probability,3(2),211-217.
[10] Singh, A. K., Singh, Priyanka and Singh, V. K. (2014b): Imputation Methods of Missing data for estimating the Population Mean using Simple Random Sampling, Global Journal of Advanced Research. Vol-1, Issue No. 2, PP. 253-263.
[11] Singh, Priyanka, Singh, A.K. and Singh, V. K. (2015): On the Use of Compromised Imputation for Missing data using Factor-Type Estimators. J. Stat. Appl. Pro. Lett. Vol-2, Issue No. 2, PP.1-9.
[12] Singh, A. K., Singh, P. and Singh, V. K. (2016a): Estimating Mean Under Non-Response in Two-Phase Sampling for Negative Correlated Data. International Journal of Mathematics and Statistics, Vol-17, No. 2,75-84.
[13] Singh, A. K., Singh, P. and Singh, V. K. (2016b): Estimation of mean under Imputation of Missing data using Exponential-Type Estimator in Two-Phase Sampling. International Journal of Statistics and Economics , Vol -17, No. 1,73-81.
[14] Singh, G. N. Maurya, S. Khetyan, M. and Kadilar, Cem (2016): Some Imputation methods for missing data in sample surveys, Hecettepe Journal of Mathematics and Statistics,45(6)1865-1880.
[15] Prasad, Shakti (2017): Ratio Exponential type estimators with imputation for missing data in sample surveys, Model Assisted Statistics and Applications, vol. 12, No. 2, PP. 95-106.

Appendix

We have

$y_{T_1} = \bar{\mathrm{y}}_{\mathrm{r}}\ T_1$, $y_{T_2} = \bar{y}_r T_2$, $y_{T_3} = \bar{\mathrm{y}}_{\mathrm{r}}\ T_3$,where T_1, T_2, T_3 are given in the expressions(17),(18),and(19) respectively. Now using the large sample approximations $\varepsilon = \frac{\bar{y}_r}{\bar{Y}} - 1,\ \delta = \frac{\bar{x}_r}{\bar{X}} - 1$ and $\eta = \frac{\bar{x}_n}{\bar{X}} - 1$.With the concept of two-phase sampling and following Rao and Sitter (1995)mechanism of MCAR, for given r and n, we have.

$$E(\varepsilon) = E(\delta) = E(\eta) = 0\ E(\varepsilon^2) = \theta_{r,N} C_Y^2;\quad E(\delta^2) = \theta_{r,N} C_X^2;\qquad E(\eta^2) = \theta_{n,N} C_X^2;$$

$$E(\varepsilon\delta) = \theta_{r,N}\rho C_Y\ C_X;\ E(\varepsilon\eta) = \theta_{n,N}\rho C_Y\ C_X;\ E(\delta\eta) = \theta_{n,N} C_X^2;$$

The estimator y_{T_1} , y_{T_2} and y_{T_3} in terms of ε, δ and η up to first order of approximation, could be expressed as:

$$y_{T_1} = \bar{\mathrm{Y}}\left[1 + \varepsilon + \alpha\left(-\frac{\eta}{2} + \frac{\eta^2}{4}\right) + \frac{\alpha^2}{8}\eta^2 - \frac{\alpha}{2}\varepsilon\eta\right] \tag{36}$$

$$y_{T_2} = \ \bar{\mathrm{Y}}\left[1 + \varepsilon + \frac{\alpha}{2}(\eta - \delta) - \frac{\alpha}{4}(\eta^2 - \delta^2) + \frac{\alpha^2}{8}(\eta^2 + \delta^2 - 2\eta\delta) + \frac{\alpha}{2}(\eta\varepsilon - \varepsilon\delta)\right] \tag{37}$$

$$y_{T_3} = \bar{Y}\left\{1 + \varepsilon - \frac{\alpha}{2}\delta + \frac{\alpha}{4}\delta^2 + \frac{\alpha^2}{8}\delta^2 - \alpha\frac{\varepsilon\delta}{2}\right\} \tag{38}$$

The expression (36), (37) and (38) obtained assuming that $|\ \varepsilon\ | < 1, |\ \eta\ | < 1and|\ \delta\ | < 1$, are valid assumptions. Taking expectation of both the sides of (36),(37)and(38) and realizing that $B(y_{T_i}) = E(y_{T_i}) - \bar{Y}, i = 1,2,3$ .we have the expressions (23),(26) and (29).

Similarly, squaring the expression (36), (37) and (38), neglecting the terms of $\varepsilon, \delta\ and\ \eta$ greater than two and realizing that

$$M\left(y_{Ti}\right) = E[y_{T_i}{}^2] + \bar{Y}^2 - 2\bar{Y}E[y_{T_i}]\ ,\qquad i = 1,2,3$$

The expressions (24), (27) and (30) could be obtained applying large sample approximation results as given above.