iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37772

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Improved Exponential Dual to Ratio Type Imputation for Missing Data under Two-Phase Sampling

B. K. Singh[1], Krishnajyothi Nath[2]
[1, 2]Departmen of Mathematicst, North Eastern Regional Institute of Science and Technology (NERIST), Itanagar, India

Abstract: In this paper, authors have proposed a class of exponential dual to ratio type compromised imputation technique and corresponding point estimator in two-phase sampling design. Two different sampling designs in two-phase sampling are compared under imputed data. The bias and M.S.E. of suggested estimator is derived in the form of population parameters using the concept of large sample approximation. Numerical study is performed over two populations using the expressions of bias and M.S.E. and efficiency compared with existing estimators.
Keywords: Missing data, Bias, Mean squared error (M.S.E), Two-phase sampling, SRSWOR, Compromised Imputation (C.I.).

I. INTRODUCTION

Missing data is a problem encountered in almost every data collection activity but particularly in sample survey. To overcome the problem of missing observations or non-response in sample surveys, the technique of imputation is frequently used to replace the missing data. In literature, several imputation techniques are described, some of them are better over others. To deal with missing values effectively Kalton et al. (1981) and Sande (1979) suggested imputation that make an incomplete data set structurally complete and its analysis simple. Lee et al. (1994, 1995) used the information on an auxiliary variate if it is available. Later Singh and Horn (2000) suggested a compromised method of imputation. Ahmed et al. (2006) discussed several new imputation based estimators that used the information on an auxiliary variate and compared their performance with the mean method of imputation. Shukla (2002) discussed F-T estimators under two-phase sampling and Shukla and Thakur (2008) have proposed estimation of mean with imputation of missing data using F-T estimators. Shukla et al. (2009) have discussed on utilization of non-response auxiliary population mean in imputation for missing observations. Shukla et al. (2009a) have further discussed on the estimation of mean under imputation for missing data using F-T estimators in two-phase sampling and further Shukla et al. (2011) have suggested linear combination based imputation methods for missing data in sample. Thakur et al. (2012) suggested some imputation methods for mean estimation in case population parameter of auxiliary information is unknown. Further Thakur et al. (2013) discussed the estimation of mean in presence of missing data under two-phase sampling scheme while the numbers of available observations are considered as random variable. The objective of the present research work is to derive some imputation methods for mean estimation in case population parameter of auxiliary information is missing or unknown.

II. NOTATIONS

Let □□□□□□1, 2... N□□be a finite population with Y_i as a variable of main interest and X_i $(i=1,2,...,N)$ an auxiliary variable. As usual, $\overline{Y} = N^{-1}\sum_{i=1}^{N} Y_i$, $\overline{X} = N^{-1}\sum_{i=1}^{N} X_i$ are population means, $\overline{X}$ is assumed unknown and $\overline{Y}$ under investigation.

Consider a preliminary large sample S' of size n' is drawn from population Ω by Simple Random Sampling (SRSWOR) and a secondary sample S of size $n(n<n')$ is drawn in either of the following manners:

Case-I: as a sub-sample from sample S' (denoted by design I) as in fig.1(a),

Case-II: independent to sample S' (denoted by design II) as in fig. 1(b), without replacing S'.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Let sample size S of n units contains r responding units $(r < n)$ forming a sub-space R and $(n-r)$ non-responding with sub-space R^C in $S = R \cup R^C$. For every $i \in R$, y_i is observed variable. For $i \in R^C$, the y_i values are missing and imputed values are computed. The i^{th} value x_i of auxiliary variate is used as a source of imputation for missing data when $i \in R^C$. Assume for S, the data $x_s = \{x_i : i \in S'\}$ and $\{x_i : i' \in S'\}$ are known with mean $\bar{x} = (n)^{-1}\sum_{i=1}^{n} x_i$ and $\bar{x}' = (n')^{-1}\sum_{i=1}^{n'} x_i$ respectively. The following symbols are used hereafter:

$\bar{X}, \bar{Y}$: the population mean of X and Y respectively;

$\bar{x}, \bar{y}$: the sample mean of X and Y respectively;

$\bar{x}_r, \bar{y}_r$: the sample mean of X and Y for corresponding responding units respectively; ρ_{xy} : the correlation co-efficient between X and Y;

S_x^2, S_y^2 : the population mean squares of X and Y respectively;

C_x, C_y : the co-efficient of variation of X and Y respectively.

$$\delta_1 = \left(\frac{1}{r} - \frac{1}{n'}\right);\ \delta_2 = \left(\frac{1}{n} - \frac{1}{n'}\right);\ \delta_3 = \left(\frac{1}{n'} - \frac{1}{N}\right);\ \delta_4 = \left(\frac{1}{r} - \frac{1}{N-n'}\right);\quad \delta_5 = \left(\frac{1}{n} - \frac{1}{N-n'}\right);\ f = \frac{n}{N}\ ; g = \frac{n}{n'-n}.$$

III. LARGE SAMPLE APPROXIMATION

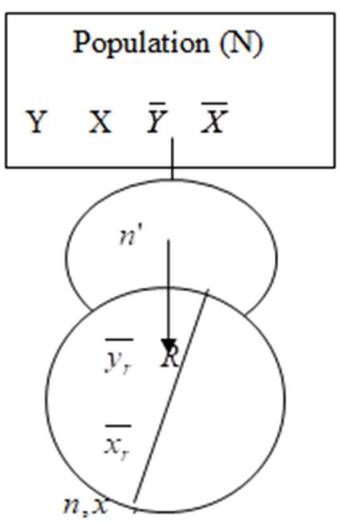

Fig.1 (a) [Design I, F$_1$]

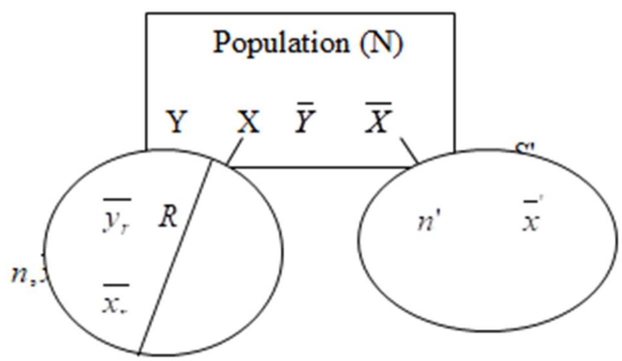

Fig.1 (b) [Design II, F$_2$]

Let $\bar{y}_r = \bar{Y}(1+e_1)$; $\bar{x}_r = \bar{X}(1+e_2)$; $\bar{x} = \bar{X}(1+e_3)$; and $\bar{x}' = \bar{X}(1+e_3')$, which implies the results $e_1 = \frac{\bar{y}_r}{\bar{Y}} - 1$;

$e_2 = \frac{\bar{x}_r}{\bar{X}} - 1$; $e_3 = \frac{\bar{x}_n}{\bar{X}} - 1$ and $e_3' = \frac{\bar{x}'}{\bar{X}} - 1$.

Now by using the concept of two-phase sampling, following Rao and Sitter (1995) and the mechanism of Missing Completely at Random (MCAR), for given r, n and n', we have:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

A. Under design F_1 [Case I]

$$E(e_1)=E(e_2)=E(e_3)=E(e_3')=0;\; E(e_1^2)=\delta_1 C_y^2;\; E(e_2^2)=\delta_1 C_x^2;\; E(e_3^2)=\delta_2 C_x^2;\; E(e_3'^2)=\delta_3 C_x^2;$$
$$E(e_1e_2)=\delta_1\rho C_y C_x;\quad E(e_1e_3)=\delta_2\rho C_y C_x;\quad E(e_1e_3')=\delta_3\rho C_y C_x;\quad E(e_2e_3)=\delta_2 C_x^2;\quad E(e_2e_3')=\delta_3 C_x^2;$$
$$E(e_3e_3')=\delta_3 C_x^2;$$

B. Under design F_2 [Case II]

$$E(e_1)=E(e_2)=E(e_3)=E(e_3')=0;\; E(e_1^2)=\delta_4 C_y^2;\; E(e_2^2)=\delta_4 C_x^2;\; E(e_3^2)=\delta_5 C_x^2;\; E(e_3'^2)=\delta_3 C_x^2;$$
$$E(e_1e_2)=\delta_4\rho C_y C_x;\; E(e_1e_3)=\delta_5\rho C_y C_x;\; E(e_1e_3')=0;\; E(e_2e_3)=\delta_5 C_x^2;\; E(e_2e_3')=0;\; E(e_3e_3')=0;$$

IV. SOME EXISTING IMPUTATION TECHNIQUE

Let $\overline{Y}=N^{-1}\sum_{i=1}^{N}y_i$ be the mean of the finite population under consideration. A Simple Random Sampling Without Replacement (SRSWOR), S of size n is drawn from $\Omega=\{1,2,...,N\}$ to estimate the population mean $\overline{Y}$. Let the number of responding units out of sampled n units be denoted by $r(r<n)$, the set of responding units, by R, and that of non-responding units by R^C. For every unit $i\in R$ the value y_i is observed, but for the units $i\in R^C$, the observations y_i are missing and instead imputed values are derived. The i^{th} value x_i of auxiliary variate is used as a source of imputation for missing data when $i\in R^C$. Assume for S, the data $x_s=\{x_i : i\in S\}$ are known with mean $\overline{x}=(n)^{-1}\sum_{i'=1}^{n}x_i$. Under this setup, some well known imputation methods are given below:

A. Mean Methods of Imputation

The mean imputation method is to replace each missing datum with the mean of the observed value. The data after imputation becomes

For y_i define y_{oi} as $y_{oi}=\begin{cases} y_i & if \quad i\in R \\ \overline{y}_r & if \quad i\in R^C \end{cases}$

Using above, the imputation-based estimators of population mean $\overline{Y}$ is $\overline{y}_m=\frac{1}{r}\sum_{i\in R}y_i=\overline{y}_r$

The bias and mean square error is given by

(i) $B(\overline{y}_m)=0$

(ii) $V(\overline{y}_m)=\left(\frac{1}{r}-\frac{1}{N}\right)S_Y^{\,2}$

B. Ratio Method of Imputation

Following the notations of Lee, et al. (1994), in the case of single imputation method, if the i^th^ unit requires imputation, the value $\hat{b}x_i$ is imputed.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

For y_i and x_i, define y_{oi} as $y_{oi} = \begin{cases} y_i & if \quad i \in R \\ \hat{b} x_i & if \quad i \in R^C \end{cases}$ where $\hat{b} = \sum_{i \in R} y_i / \sum_{i \in R} x_i$

Using above, the imputation-based estimator is: $\bar{y}_s = \frac{1}{n} \sum_{i \in S} y_{oi} = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right) = \bar{y}_{RAT}$

where $\bar{y}_r = \frac{1}{r} \sum_{i \in R} y_i$, $\bar{x}_r = \frac{1}{r} \sum_{i \in R} x_i$ and $\bar{x}_n = \frac{1}{n} \sum_{i \in S} x_i$

The bias and mean square error of $\bar{y}_{RAT}$ is given by

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

(ii) $M(\bar{y}_{RAT}) = \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) \left[S_Y^2 + R_1^2 S_X^2 - 2 R_1 S_{XY} \right]$ where $R_1 = \frac{\bar{Y}}{\bar{X}}$

C. Compromised Method of imputation

Singh and Horn (2000) suggested a compromised method of imputation. It based on using information from imputed values for the responding units in addition to non-responding units. In case of compromised imputation procedures, the data take the form

$$y_{oi} = \begin{cases} (\alpha n / r) y_i + (1 - \alpha) \hat{b} x_i & if \quad i \in R \\ (1 - \alpha) \hat{b} x_i & if \quad i \in R^C \end{cases}$$

where α is a suitably chosen constant, such that the resultant variance of the estimator is optimum. The imputation-based estimator, for this case, is

$$\bar{y}_{COMP} = \left[\alpha \bar{y}_r + (1 - \alpha) \bar{y}_r \frac{\bar{x}}{\bar{x}_r} \right]$$

The bias, mean square error and minimum mean square error at $\alpha = 1 - \rho \frac{C_Y}{C_X}$ of $\bar{y}_{COMP}$ are given by

(i) $B(\bar{y}_{COMP}) = \bar{Y}(1 - \alpha) \left(\frac{1}{r} - \frac{1}{n} \right) (C_X^2 - \rho C_Y C_X)$

(ii) $M(\bar{y}_{COMP}) = \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) \left[S_Y^2 + R_1^2 - 2 R_1 S_{XY} \right] \right\} - \left(\frac{1}{r} - \frac{1}{n} \right) \alpha^2 \bar{Y}^2 C_X^2$

(iii) $M(\bar{y}_{COMP})_{\min} = \left[\left(\frac{1}{r} - \frac{1}{N} \right) - \left(\frac{1}{r} - \frac{1}{n} \right) \rho^2 \right] S_Y^2$

V. PROPOSED METHOD OF IMPUTATION AND ITS ESTIMATOR

Let y_i denotes the i^{th} observation of the suggested imputation strategy. We suggest the following imputation method:

$$(1)\ y_{edrdi} = \begin{cases} \theta \dfrac{n}{r} y_i + (1-\theta)\bar{y}_r \exp\left[\dfrac{\psi - \bar{x}'}{\psi + \bar{x}'}\right] & if \quad i \in R \\ (1-\theta)\bar{y}_r \exp\left[\dfrac{\psi - \bar{x}'}{\psi + \bar{x}'}\right] & if \quad i \in R^C \end{cases}$$

where $\psi = \dfrac{n'\bar{x}' - n\bar{x}_r}{n' - n}$

Under this strategy the point estimator of $\bar{Y}$ is

$$T_{edrd} = \theta \bar{y}_r + (1-\theta)\bar{y}_r \exp\left[\frac{\psi - \bar{x}'}{\psi + \bar{x}'}\right] \qquad (1)$$

where θ is a suitably chosen constant, such that the resultant variance of the estimator is minimum.

VI. PROPERTIES OF PROPOSED ESTIMATOR

Let $B(.)_t$ and $M(.)_t$ denote the bias and mean square error (M.S.E.) of the estimator under the given sampling design $t = I, II$. The properties of T_{edrd} are derived in the following theorem respectively.

A. *Theorem 6.1*

Estimator T_{edrd} in terms of e_i; i=1, 2, 3 and e_3' could be expressed upto first order of approximation:

$$T_{edrd} = \bar{Y}\left[1 + e_1 + (1-\theta)\left\{\frac{g}{2}\left(e_3' - e_2 + e_1 e_3 - e_1 e_2\right) + \frac{g^2}{8}\left(2e_2 e_3' - e_3'^2 - e_2^2\right) + \frac{g}{2}\left(e_1 e_3' - e_1 e_2\right)\right\}\right]$$

by ignoring the terms $E\left[e_i^r, e_j^s\right]$, $E\left[e_i^r, (e_j')^s\right]$ for $r + s > 2$, where $r.s = 0,1,2,...$ and $i = 1,2,3;\ j = 2,3$ which is first approximation.

Proof: $T_{edrd} = \theta \bar{y}_r + (1-\theta)\bar{y}_r \exp\left[\dfrac{\psi - \bar{x}'}{\psi + \bar{x}'}\right]$

$$= \bar{Y}\left[\theta(1+e_1) + (1-\theta)(1+e_1)\exp\left\{\frac{g(e_3' - e_2)}{2}\left(1 + \frac{2e_3' + g(e_3' - e_2)}{2}\right)^{-1}\right\}\right]$$

$$= \bar{Y}\left[1 + e_1 + (1-\theta)\left\{\frac{g}{2}\left(e_3' - e_2 + e_1 e_3 - e_1 e_2\right) + \frac{g^2}{8}\left(2e_2 e_3' - e_3'^2 - e_2^2\right) + \frac{g}{2}\left(e_1 e_3' - e_1 e_2\right)\right\}\right] \qquad (2)$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

B. Theorem 6.2

Bias of T_{edrd} under design I and II, up to first order of approximation is:

(i) $$B(T_{edrd})_I = -\bar{Y}\left[(1-\theta)\left\{(\delta_1-\delta_3)\frac{g}{2}\left(\frac{g}{4}C_X+\rho C_Y\right)C_X\right\}\right]$$

(ii) $$B(T_{edrd})_{II} = -\bar{Y}\left[(1-\theta)\left\{\left(\frac{g}{2}+\frac{g^2}{8}\right)\delta_3+\left(\frac{g^2}{8}+\frac{g}{2}\rho\frac{C_Y}{C_X}\right)\delta_4 C_X^2\right\}\right]$$

Proof: (i) $B(T_{edrd})_I = E\left[T_{edrd}-\bar{Y}\right]_I$

$$=\bar{Y}E\left[1+e_1+(1-\theta)\left\{\frac{g}{2}\left(e_3^{'}-e_2+e_1e_3^{'}-e_1e_2\right)+\frac{g^2}{8}\left(2e_2e_3^{'}-e_3^{'2}-e_2^2\right)+\frac{g}{2}\left(e_1e_3^{'}-e_1e_2\right)\right\}-1\right]$$

Using the results of design F_1 [Case I], we get the bias of the estimator

$$B(T_{edrd})_I = -\bar{Y}\left[(1-\theta)\left\{(\delta_1-\delta_3)\frac{g}{2}\left(\frac{g}{4}+\rho\frac{C_Y}{C_X}\right)C_X^2\right\}\right]$$

(ii) $B(T_{edrd})_{II} = E\left[T_{edrd}-\bar{Y}\right]_{II}$

$$=\bar{Y}E\left[1+e_1+(1-\theta)\left\{\frac{g}{2}\left(e_3^{'}-e_2+e_1e_3^{'}-e_1e_2\right)+\frac{g^2}{8}\left(2e_2e_3^{'}-e_3^{'2}-e_2^2\right)+\frac{g}{2}\left(e_1e_3^{'}-e_1e_2\right)\right\}-1\right]$$

Using the results of design F_2 [Case II], we get the bias of the estimator

$$B(T_{edrd})_{II} = -\bar{Y}\left[(1-\theta)\left\{\left(\frac{g}{2}+\frac{g^2}{8}\right)\delta_3+\left(\frac{g^2}{8}+\frac{g}{2}\rho\frac{C_Y}{C_X}\right)\delta_4 C_X^2\right\}\right]$$

C. Theorem 6.3

Mean squared error of T_{edrd} under the design I and II, upto first order of approximation can be written as:

1) $$M(T_{edrd})_I = \bar{Y}^2\left[\delta_1 C_y^2+(\delta_1-\delta_2)\left\{\frac{g^2}{4}(1-\theta)^2C_X^2-(1-\theta)g\rho C_Y C_X\right\}\right] \quad (3)$$

2) $$M(T_{edrd})_{II} = \bar{Y}^2\left[\delta_4 C_y^2+(\delta_3+\delta_4)\left\{\frac{g^2}{4}(1-\theta)^2\right\}C_X^2-(1-\theta)g\delta_4\rho C_Y C_X\right] \quad (4)$$

Proof: Squaring and taking expectations on both the sides of (2) and neglecting second and higher order terms, we get the MSE of T_{edrd} to the first degree of approximation as

a) $$M(T_{edrd})_I = E\left[T_{edrd}-\bar{Y}\right]_I^2 = \bar{Y}^2 E\left[e_1+(1-\theta)\frac{g}{2}(e_3^{'}-e_2)\right]^2$$

$$=\bar{Y}^2 E\left[e_1^2+(1-\theta)^2\frac{g^2}{4}\left(e_3^{'2}+e_2^2-2e_2e_3^{'}\right)+(1-\theta)g\left(e_1e_3^{'}-e_1e_2\right)\right] \quad (5)$$

Using the results of *design F_1 [Case 'I]*, in equation (5), we get the mean square error (MSE) of the estimator

$$M(T_{edrd})_I = \bar{Y}^2\left[\delta_1 C_y^2 + (\delta_1 - \delta_2)\left\{\frac{g^2}{4}(1-\theta)^2 C_X^2 - (1-\theta)g\rho C_Y C_X\right\}\right]$$

b) $M(T_{edrd})_{II} = E\left[T_{edrd} - \bar{Y}\right]_{II}^2 = \bar{Y}^2 E\left[e_1 + (1-\theta)\frac{g}{2}(e_3' - e_2)\right]^2$

Using the results of design F_2 [Case II], in equation (5), we get the mean square error (MSE) of the estimator

$$M(T_{edrd})_{II} = \bar{Y}^2\left[\delta_4 C_y^2 + (\delta_3 + \delta_4)\left\{\frac{g^2}{4}(1-\theta)^2\right\}C_X^2 - (1-\theta)g\delta_4\rho C_Y C_X\right]$$

D. Theorem 6.4

Minimum mean squared error of T_{edrd} under design *I* and *II* is:

1) $\left[M(T_{edrd})_I\right]_{\min} = \left[(1-\rho^2)\delta_1 + \delta_3\rho^2\right]S_Y^2$

$$\text{when } \theta = 1 - \frac{2}{g}\rho\frac{C_Y}{C_X} \qquad (4)$$

2) $\left[M(T_{edrd})_{II}\right]_{\min} = \left[\{\delta_3 + \delta_4(1-\rho^2)\}(\delta_3+\delta_4)^{-1}\delta_4\right]S_Y^2$

$$\text{when } \theta = 1 - \frac{2}{g}\frac{\delta_4}{(\delta_3+\delta_4)}\rho\frac{C_Y}{C_X} \qquad (5)$$

Proof: (i) Differentiate (3) with respect to θ and equating to zero, we get

$$\frac{d}{d\theta}\left[M(T_{edrd})_I\right] = 0 \Rightarrow \theta = 1 - \frac{2}{g}\rho\frac{C_Y}{C_X}$$

putting the value of θ in (3), we obtain

$$\left[M(T_{edrd})_I\right]_{\min} = \left[(1-\rho^2)\delta_1 + \delta_3\rho^2\right]S_Y^2$$

(ii) Similarly, proceeding for (4), we have

$$\frac{d}{d\theta}\left[M(T_{edrd})_{II}\right] = 0 \Rightarrow \theta = 1 - \frac{2}{g}\frac{\delta_4}{(\delta_3+\delta_4)}\rho\frac{C_Y}{C_X}$$

putting the value of θ in (4), we obtain

$$\left[M(T_{edrd})_{II}\right]_{\min} = \left[\{\delta_3 + \delta_4(1-\rho^2\}(\delta_3+\delta_4)^{-1}\delta_4\right]S_Y^2$$

VII. COMPARISONS OF THE ESTIMATOR

1) Comparison of the estimator $\left[M(T_{edrd})_I\right]_{\min}$ and the estimator $V(\bar{y}_s)$

$$\Delta_1 = V(\bar{y}_s) - \left[M(T_{edrd})_I\right]_{\min} = \left(\frac{1}{r} - \frac{1}{n'}\right)S_Y^2 + \left(\frac{1}{r} - \frac{2}{n'} + \frac{1}{N}\right)\rho^2 S_Y^2$$

$(T_{edrd})_I$ is better than $\bar{y}_s$ if $\Delta_1 > 0 \Rightarrow r < \dfrac{Nn'}{2N - n'}$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

2) Comparison of the estimator $\left[M\left(T_{edrd}\right)_{II}\right]_{\min}$ and the estimator $V(\bar{y}_s)$

$$\Delta_2 = \mathrm{V}\left(\bar{\mathrm{y}}_s\right) - \left[M\left(T_{edrd}\right)_{II}\right]_{\min} = \left(\frac{1}{N-n'} - \frac{1}{N}\right)S_Y^2 + \frac{\left(\frac{1}{r} - \frac{1}{N-n'}\right)^2 \rho^2 S_Y^2}{\left(\frac{1}{n'} - \frac{1}{N} + \frac{1}{r} - \frac{1}{N-n'}\right)}$$

$\left(T_{edrd}\right)_{II}$ is better than $\bar{y}_s$ if $\Delta_2 > 0 \Rightarrow r < \dfrac{N\left(N-n'\right)n'}{3Nn' - n'^2 - N^2}$

3) Comparison of the estimator $\left[M\left(T_{edrd}\right)_{I}\right]_{\min}$ and the estimator $M(\bar{y}_{RAT})$

$$\Delta_3 = \mathrm{M}\left(\bar{\mathrm{y}}_{RAT}\right) - \left[M\left(T_{edrd}\right)_{I}\right]_{\min}$$
$$= \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)C_Y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_X^2 - 2\rho C_Y C_X\right) + \left(\frac{1}{r} - \frac{2}{n'} + \frac{1}{N}\right)\rho^2 C_Y^2\right]$$

$\left(T_{edrd}\right)_{I}$ is better than $\bar{y}_{RAT}$, if $\Delta_3 > 0$.

this generates two conditions as

(i) when $\left(C_X^2 - 2\rho C_Y C_X\right) > 0 \Rightarrow \rho \dfrac{C_Y}{C_X} < \dfrac{1}{2}$

(ii) when $\left(\dfrac{1}{r} - \dfrac{2}{n'} + \dfrac{1}{N}\right) > 0 \Rightarrow r < \dfrac{Nn'}{2N - n'}$

4) Comparison of the estimator $\left[M\left(T_{edrd}\right)_{II}\right]_{\min}$ and the estimator $M(\bar{y}_{RAT})$

$$\Delta_4 = \mathrm{M}\left(\bar{\mathrm{y}}_{RAT}\right) - \left[M\left(T_{edrd}\right)_{II}\right]_{\min}$$
$$= \bar{Y}^2\left[\left(\frac{1}{N-n'} - \frac{1}{n}\right)C_Y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_X^2 - 2\rho C_Y C_X\right) + \left\{\frac{\left(\frac{1}{r} - \frac{1}{N-n'}\right)^2}{\left(\frac{1}{n'} - \frac{1}{N}\right) + \left(\frac{1}{r} - \frac{1}{N-n'}\right)}\right\}\rho^2 C_Y^2\right] > 0$$

$\left(T_{edrd}\right)_{II}$ is better than $\bar{y}_{RAT}$ if $\Delta_4 > 0$

this generates two conditions as

(i) when $\left(C_X^2 - 2\rho C_Y C_X\right) > 0 \Rightarrow \rho \dfrac{C_Y}{C_X} < \dfrac{1}{2}$

(ii) when $\left(\dfrac{1}{n'} - \dfrac{1}{N} + \dfrac{1}{r} - \dfrac{1}{N-n'}\right) > 0 \Rightarrow r < \dfrac{N\left(N-n'\right)n'}{3Nn' - n'^2 - N^2}$

5) Comparison of the estimator $\left[M\left(T_{edrd}\right)_I\right]_{\min}$ and the estimator $M\left(\bar{y}_{COMP}\right)$

$$\Delta_5 = \mathrm{M}\left(\bar{\mathrm{y}}_{COMP}\right) - \left[M\left(T_{edrd}\right)_I\right]_{\min}$$

$$= \bar{Y}^2\left[\left(\frac{1}{n'}-\frac{1}{N}\right)C_Y^2 - \left(\frac{2}{n'}-\frac{1}{n}-\frac{1}{N}\right)\rho^2 C_Y^2\right] > 0$$

$\left(T_{edrd}\right)_I$ is better than $\bar{y}_{COMP}$, if $\Delta_5 > 0 \Rightarrow \left(\frac{2}{n'}-\frac{1}{n}-\frac{1}{N}\right) < 0$

$$\Rightarrow n < \frac{Nn'}{2N-n.}$$

6) Comparison of the estimator $\left[M\left(T_{edrd}\right)_{II}\right]_{\min}$ and the estimator $M\left(\bar{y}_{COMP}\right)$

$$\Delta_6 = \mathrm{M}\left(\bar{\mathrm{y}}_{COMP}\right) - \left[M\left(T_{edrd}\right)_{II}\right]_{\min}$$

$$= \left(\delta_6 - \delta_4\right)S_Y^2 - \left(\frac{\delta_7\left(\delta_3+\delta_4\right)-\delta_4^2}{\left(\delta_3+\delta_4\right)}\right)\rho^2 S_Y^2$$

$\left(T_{edrd}\right)_{II}$ is better than $\bar{y}_{COMP}$, if $\Delta_6 > 0$

$$\Rightarrow \rho^2 < \frac{\left(\delta_6-\delta_4\right)\left(\delta_3+\delta_4\right)}{\left[\delta_7\left(\delta_3+\delta_4\right)-\delta_4^2\right]} \Rightarrow -\sqrt{P} < \rho < \sqrt{P}$$

where, $P = \frac{\left(\delta_6-\delta_4\right)\left(\delta_3+\delta_4\right)}{\left[\delta_7\left(\delta_3+\delta_4\right)-\delta_4^2\right]}$; $\delta_6 = \left(\frac{1}{r}-\frac{1}{N}\right)$; $\delta_7 = \left(\frac{1}{r}-\frac{1}{n}\right)$.

VIII. NUMERICAL ILLUSTRATIONS

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

Table 8.1: Parameters of Population A and B

Population	N	$\bar{Y}$	$\bar{X}$	S_y^2	S_x^2	ρ	C_x	C_y
A	200	42.48518	18.515	199.0598	48.5375	0.8652	0.3763	0.3321
B	8306	253.75	343.316	338006	862017	0.522231	2.70436	2.29116

Let $n' = 60, n = 40, r = 5, g = 2$ for population A and $n' = 2000$, $n = 500$, $r = 450$, $g = 0.33333$ for population B respectively.

The percent relative efficiency of different estimators is shown in tables 8.2 and 8.3.

Table 8.2: MSE, bias, and percent relative efficiencies (PRE) of different estimators for Population A

Estimators	Population A		
	Bias	Efficiency	MSE
$\bar{y}_s$	0	100	38.81893
$\bar{y}_{RAT}$	0.24890	254.70485	15.24075
$\bar{y}_{COMP}$	0.19005	304.68706	12.74059
T_{edrd}	-0.99641	355.67436	10.91417

Table 8.3: MSE, bias, and percent relative efficiencies (PRE) of different estimators for Population B

Estimators	Population B		
	Bias	Efficiency	MSE
$\bar{y}_s$	0	100	710.4302
$\bar{y}_{RAT}$	0.22994	92.3546	768.7752
$\bar{y}_{COMP}$	0.05041	102.9321	689.9429
T_{edrd}	-27.91688	154.9022	458.3537

IX. CONCLUSION

In this paper, the PRE of the suggested estimator T_{edrd} has been compared with several other estimators, viz., $\bar{y}_s, \bar{y}_{RAT}$, *and* $\bar{y}_{COMP}$. From tables 8.2 and 8.3, it is observed that the proposed estimator T_{edrd} in its optimality is more efficient than the other estimators taken for comparisons under considerations. Hence, the proposed estimator is preferable in comparison to other estimators taken into consideration.

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, Statsitics in Transition, 7(6), p. 1247-1264.
[2] Kalton, G., Kasprzyk, D. and Santos, R. (1981). Issues of non-response and imputation of income and participation. Current topics in survey sampling. In D. Krevoski, R. Platek and J N. K. Rao(Eds.), (pp.455-480). New York: Acad. Press.
[3] Lee, H., Rancourt, E. and Sarndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. Journal of Official Statistics, 10(3). P. 231-243.
[4] Lee, H., Rancourt, E. and Sarndal, C. E. (1995). Variance estimation in the presence of imputed data for the generalized estimation system. Proc. of the American Statist. Assoc. (Social Survey Research Methods Section), p. 384-389.
[5] Sande, I. G. (1979). A personal view of hot deck approach to automatic edit and imputation. Journal Imputation Procedures. Survey Methodology, 5, 238-246.
[6] Singh, A. K., Singh, P. and Singh, V. K. (2014). Imputation methods of missing data for estimating the population mean using simple random sampling, Global Journal of Advance Research, 1(2), p.253-263.
[7] Shukla, D. (2002): F-T estimator under two-phase sampling, Metron, 59, 1-2, p. 253-263.
[8] Shukla, D., and Thakur, N. S. (2008). Estimation of mean with imputation of missing data using factor-type estimator, Statistics in Transition, 9(1), p. 33-48.
[9] Shukla, D., Thakur, N. S., Pathak, S. and Rajput D. S. (2009). Estimation of mean with imputation of missing data using factor-type estimator in two-phase sampling, Statistics in Transition, 10(3), p. 397-414.
[10] Shukla, D., Thakur, N. S., Thakur, D. S. (2009). Utilization of non-response auxiliary population mean in imputation for missing observations, Journal of Reliability and Statistical Studies, 2(1), p. 28-40.

[11] Shukla, D., Thakur, N. S., Thakur, D. S. and Pathak, S. (2011). Linear combination based imputation method for missing data in sample, International Journal of Modern Engineering Research (IJMER), 1(2). p. 580-596.
[12] Singh S. and Horn S. (2000). Compromised imputation in survey sampling, Metrika, 51, p. 266-276.
[13] Thakur, N. S. and Yadav, K. (2011). Estimation of mean in presence of missing data under two-phase sampling scheme, Journal of Reliability and Statistical Studies, 4(2), p. 93-104.
[14] Thakur, N. S., Yadav, K. and Pathak, S. (2013). On mean estimation with imputation in two phase sampling design, Research Journal of Mathematical and Statistical Sciences, 1(3),p.1-9.