*International Journal of Basic and Applied Science 12 (4) (2024) 148-158*

Published by: IOCSCIENCE

International Journal of Basic and Applied Science

Journal homepage: www.ijobas.pelnus.ac.id

# Ratio estimator for double sampling procedure with non-response: An empirical study

**Oguagbaka, S. K.[1], Okoli, O. C.[2] and Aronu, C. O. [3]**

[1]Department of Statistics, Federal Polytechnic, Oko, Anambra State, Nigeria

[2]Department of Mathematics, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria

[3]Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria

**Article Info**

***Article history:***

Received: Dec 27, 2023
Revised: Feb 04, 2024
Accepted: Mar 26, 2024

***Keywords:***

Auxiliary variables;
Double sampling;
Non-response;
Population mean;
Ratio-type estimators,

**ABSTRACT**

This study proposes a ratio-type estimator for population mean estimation using auxiliary variables with double sampling in the presence of non-response. The study provides expressions for the constant, bias, and mean square errors (MSE) of the proposed estimator and compares it with ten existing estimators. The study employed the secondary source of data collection to evaluate the efficiency of the proposed and existing estimators by analyzing five natural populations from three different sources. The performance of ten (10) estimators was considered in this study. The findings suggest that the proposed estimator and the H estimator provide more accurate and precise estimates of the population mean using an auxiliary variable. Additionally, the study found significant differences amongst the mean values of the constant and bias for the different estimators. A Dunn Kruskal-Wallis multiple comparison tests with the Bonferroni method was performed to ascertain the pair of estimators that contributed to the significant difference observed. When estimating the population means using an auxiliary variable, the proposed estimator outperformed other existing estimators that were taken into consideration in the study.



***Corresponding Author:***

Aronu, C. O.,
Department of Statistics,
Chukwuemeka Odumegwu Ojukwu University,
QR9P+WFR, Uli, Atughobi, Nigeria,
Email: amaro4baya@yahoo.com

## 1. INTRODUCTION

Survey sampling is a widely used technique for estimating population parameters using data from a sample. One of the most commonly used approaches for estimating population parameters is the ratio-type estimator, which utilizes information on auxiliary variables to improve the accuracy of estimation. Double sampling has also been widely used in situations where the cost of measuring the variable of interest is high or when a portion of the sample is lost due to non-response. However, non-response in double sampling poses a challenge in estimating the population mean, and there is a need for new and improved methods.

Various researchers have proposed different methods to address non-response issues in survey sampling, particularly in double sampling. The study by [1] proposed a modified ratio-product exponential estimator to estimate the population mean of the study variable in the presence of non-response. Also, the study by [2] had suggested a similar estimator. The authors showed that the proposed estimator was more effective than the typical unbiased estimators, ratio estimators, product estimators, and exponential ratio and product estimators published by [3].

According to the study by [4], non-response impacts almost all surveys. It was suggested to use a postal questionnaire for the first effort and a personal interview for the second attempt, assuming that a subsample of the initial non-respondents would be reached again using a more costly method. In a study by [5], double sampling and auxiliary data were used to evaluate the challenge of calculating the population mean when there is no response. In the event of a non-response, they provided a general class of estimators for the population mean of the research variable and discovered that the proposed estimators outperform the optimal estimator. In his contribution, [6] proposed an improved approach for estimating the population mean of the study variable by using known values for a particular population characteristic or variables, along with additional data. A class of estimators based on data from two auxiliary variables was reported in the study by [7], which discovered that the suggested class of estimators consistently outperformed the regression estimators based on two variables. The work of [8] highlighted the importance of evaluating the finite population variance in several domains, including health. They suggested estimating the population variance using the population mean of an auxiliary variable and a generalised form of exponential-type estimators. The proposed generalised estimator was found to be more efficient than the conventional unbiased estimator, the conventional ratio and product, the exponential-type estimators, and other estimators in the same class by the authors after they compared and looked into a few particular applications of the proposed estimator.

Theorised by [9, 10] are exponential ratio type estimators that account for missing data on the auxiliary variable, the research variable, or both. Regression, product, and ratio estimators are used in a novel way by [11] to address non-response issues in double sampling. According to a similar study [12], an exponential ratio type estimator outperformed traditional unbiased estimators to estimate the population mean using auxiliary data, even in the face of measurement errors and non-response. This research offered important new perspectives on how to address missing data in survey sample scenarios so that population mean estimators perform more accurately. Moreover, the work by [13] introduced a unique class of exponential-type estimators for estimating the finite population total of the variable under study. It showed the improved efficiency of their proposed estimators over existing estimators. To estimate the finite population mean of the study variable, a novel weighted estimator was presented by [14], and it was demonstrated to be more effective than previous estimators. However, the work by [15] showed that it was better than previous estimators in estimating the population mean of the research variable in the context of non-response using known values of an auxiliary variable. The studies reviewed emphasise how crucial it is to manage non-response in sample surveys and how useful it is to use auxiliary data to raise estimator accuracy.

In contrast, a significant gap is anticipated to be closed by the current study about survey sample techniques and non-response. The present research proposes a ratio-type estimator for population mean estimation using auxiliary variables with double sampling in the presence of non-response. This was achieved by deriving the expressions for the constant, bias, and mean square errors (MSE) of the proposed estimator. The proposed estimator was compared to ten other existing ratio-type estimators, including Sisodia and Dwivedi (SD) [16], Upadhyaya and Singh (US) [17], Singh and Tailor (ST) [18], Singh et al. (SET) [19], Yan and Tain (YT) [20], Subramani and Krumarpandiyan (SK) [21], Hazara (H) [22], Jerajuddin and Kishun (JK) [23], Ijaz et al. (IET) [24], and Suleiman and Adewara (SA) [25]. Three distinct sources of five natural populations were used by the present study to assess the estimators' performance. Moreover, the outcome of this research are expected to have a significant impact on future research in this area and contribute to the development of more accurate and practical survey sampling methods.

## 2. RESEARCH METHOD

### 2.1. Method of Data Collection

The principal source of information for the study was secondary data that was gathered from published journals or already made data. To meet the objectives of the study, this method of data collection comprised identifying and collecting pertinent data from the body of scholarly literature already in existence. By using published publications, a thorough and reliable basis for the data analysis and interpretation of the study was guaranteed.

### 2.2. Method of Data Analysis

The study follows the approach proposed by [22] and extends it by combining two existing ratio-type estimators to improve the accuracy of estimation. When measuring the variable of interest is expensive or when non-response causes a portion of the sample to be lost, the proposed approach may be useful. Suppose that the study variable's sample mean based on $n_1$ and $r$ units is $(\bar{y}_1, \bar{y}_2)$ and that the auxiliary variable's sample mean based on $n_1, r, and\ n$ units is $(\bar{x}_1, \bar{x}_2', \bar{x}', \bar{x})$. Where $\bar{Y}$ and $\bar{X}$ represent the research variable's population mean and the auxiliary variable, respectively, depending on the population size $N = N_1 + N_2$. Furthermore, the population means of the research variable, $\bar{Y}_2$ and $\bar{X}_2$, and the auxiliary variable, $M_d$ (median), $C_{y,}C_x$ (Coefficient of variation), and $N_2$(non-response part) are dependent on the population.

**Nomenclature:**

$S_y^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{(N-1)}$ , (the finite population varinace of Y)

$S_x^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{(N-1)}$ , (the finite population varinace of X)

$S_{y2}^2 = \frac{\sum_{i=1}^{N_2}(y_i - \bar{Y}_2)^2}{(N_2-1)}$ , (the finite population varinace of $Y_2$)

$S_{x2}^2 = \frac{\sum_{i=1}^{N_2}(x_i - \bar{X}_2)^2}{(N_2-1)}$ , (the finite population varinace of $X_2$)

$C_y = \frac{S_y}{\bar{Y}}$ , (coefficient of variation of Y)

$C_{y2} = \frac{S_{y2}}{\bar{Y}}$ , (coefficient of variation of $Y_2$)

$C_x^2 = \frac{S_x^2}{\bar{X}^2}$, (coefficient of variation of $Y_2$)

$C_{x2}^2 = \frac{S_{x2}^2}{\bar{X}^2}$, (coefficient of variation of $X_2$)

$S_{yx} = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})}{(N-1)}$ , (the finite population covarinace of X and Y)

$S_{yx2} = \frac{\sum_{i=1}^{N_2}(y_i - \bar{Y})(x_i - \bar{X}_2)}{(N_2-1)}$ , (the finite population covarinace of $X_2$ and Y)

$\rho_{yx} = \frac{S_{yx}}{S_y S_x}$, Pearson's Moment Correlation Coefficient of X and Y)

$\rho_{yx2} = \frac{S_{yx2}}{S_{y2} S_{x2}}$, (Pearson's Moment Correlation Coefficient of $X_2$ and Y)

$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$ , (the finite population varinace of the auxiliary variable $x$)

$\beta_{1(x)} = \frac{N\sum_{i=1}^{N}(X_i - \bar{X})^3}{(N-1)(N-2)S_x^3}$ , (the coefficient of Skewness of the auxiliary variable $x$)

$\beta_{2(x)} = \frac{N(N+1)\sum_{i=1}^{N}(X_i - \bar{X})^4}{(N-1)(N-2)S_x^4} - \frac{3(N-1)^2}{(N-2)(N-3)}$ , (the coefficient of Kurtosis of the auxiliary variable $x$)

$M_d$ =Median of the auxiliary variable X.

Where, based on $n_1 + r$ units, $\hat{S}_{yx}$ and $\hat{S}_x^2$ are estimations for $S_{yx}$ and $S_x^2$, respectively. However, the correlation coefficients of the response and non-response groups between the study variable and the auxiliary variable x are, respectively, $\rho_{yx}$ and $\rho_{yx2}$.

**The Proposed Modified Estimator with Bias and MSE**

To estimate the population mean utilizing auxiliary variables with double sampling in the event of non-response, we can suggest a novel two ratio type estimator, as shown in [22]:

$$T_p = \bar{y}^*\left(\alpha\left(\frac{\bar{x}^*}{\bar{x}'}\right) + (1-\alpha)\left(\frac{\bar{x}}{\bar{x}'}\right)\right) \quad (1)$$

Where $\alpha$ is a constant

To obtain the bias and variance of the estimator $T_p$, we can write

$\bar{y}^* = \bar{Y}(1-\varepsilon_0), \bar{x}^* = \bar{X}(1-\varepsilon_1), \bar{x}' = \bar{X}(1-\varepsilon_1), and\ \bar{x} = \bar{X}(1-\varepsilon_2)$

Such that

$E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_1') = E(\varepsilon_2) = 0$

and

$E(\varepsilon_0^2) = \lambda C_y^2 + \lambda^* C_{y2}^2, E(\varepsilon_1^2) = \lambda C_x^2 + \lambda^* C_{x2}^2, E(\varepsilon_1'^2) = \lambda' C_x^2, E(\varepsilon_2^2) = \lambda C_x^2,\ E(\varepsilon_0\varepsilon_1) = \lambda\rho_{yx}C_xC_y + \lambda^*\rho_{yx2}C_{y2}C_{x2},$

$E(\varepsilon_0\varepsilon_2) = \lambda\rho_{yx}C_xC_y, E(\varepsilon_1\varepsilon_1') = \lambda' C_x^2, E(\varepsilon_1\varepsilon_2) = \lambda C_x^2, E(\varepsilon_2\varepsilon_1') = \lambda' C_x^2, and\ E(\varepsilon_0\varepsilon_1') = \lambda'\rho_{yx}C_xC_y.$

The estimator can be expressed in terms of ε's as follows:

$$T_p = \bar{Y}(1+\varepsilon_0)\left(\alpha\left((1+\varepsilon_0)(1+\varepsilon_1')^{-1}\right) + (1-\alpha)\left((1+\varepsilon_2)(1+\varepsilon_1')^{-1}\right)\right) \quad (2)$$

Given the assumptions that $|\varepsilon_0| < 1, |\varepsilon_1'| < 1, |\varepsilon_1| < 1, and\ |\varepsilon_2| < 1$ then the right hand side of equation (2) can be expanded to second degree of approximation. We can then have

$$T_p = \bar{Y}(1+\varepsilon_0)\left(\alpha\left(1-\varepsilon_1' + \frac{(\varepsilon_1')^2}{2} + \varepsilon_1 - \varepsilon_1\varepsilon_1' + \frac{\varepsilon_1(\varepsilon_1')^2}{2}\right) + (1-\alpha)\left(1-\varepsilon_1' + \frac{(\varepsilon_1')^2}{2} + \varepsilon_2 - \varepsilon_2\varepsilon_1' + \frac{\varepsilon_2(\varepsilon_1')^2}{2}\right)\right) \quad (3)$$

$$T_p = \bar{Y}(1+\varepsilon_0)\left(\alpha\left(1+\varepsilon_1-\varepsilon_1'-\varepsilon_1\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right) + (1-\alpha)\left(1+\varepsilon_2-\varepsilon_1'-\varepsilon_2\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right)$$

$$T_p - \bar{Y} = \bar{Y}(1+\varepsilon_0)\left(\alpha\left(1+\varepsilon_1-\varepsilon_1'-\varepsilon_1\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right) + (1-\alpha)\left(1+\varepsilon_2-\varepsilon_1'-\varepsilon_2\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right) - \bar{Y}$$

$$T_p = \bar{Y}\left(\alpha\left(1+\varepsilon_0+\varepsilon_1-\varepsilon_1'-\varepsilon_1\varepsilon_1'+\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right) + (1-\alpha)\left(1+\varepsilon_0+\varepsilon_2-\varepsilon_1'-\varepsilon_2\varepsilon_1'-\varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right) - \bar{Y}$$

$$T_p = \bar{Y}\left(\alpha\left(\varepsilon_0+\varepsilon_1-\varepsilon_1'-\varepsilon_1\varepsilon_1'+\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right) + (1-\alpha)\left(\varepsilon_0+\varepsilon_2-\varepsilon_1'-\varepsilon_2\varepsilon_1'-\varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right) \quad (4)$$

Taking expectation on both sides of equation (4), we shall obtain the bias of the $T_p$ as

$$Bias(T_p) = \bar{Y}\left(\alpha E\left(\varepsilon_0+\varepsilon_1-\varepsilon_1'-\varepsilon_1\varepsilon_1'+\varepsilon_0\varepsilon_1-\varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right) + (1-\alpha)E\left(\varepsilon_0+\varepsilon_2-\varepsilon_1'-\varepsilon_2\varepsilon_1'-\varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right) \quad (5)$$

$$T_p = \overline{Y}\left(\alpha\left(-\lambda C_x^2 + \lambda\rho_{yx}C_xC_y - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right) + (1-\alpha)\left(-\lambda C_x^2 - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right)\right)$$

$$= \overline{Y}\left(\alpha\left(-\lambda C_x^2 + \lambda\rho_{yx}C_xC_y - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right) - \alpha\left(-\lambda C_x^2 - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right) + \left(-\lambda C_x^2 - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right)\right)$$

$$= \overline{Y}\left(\alpha\left(\lambda\rho_{yx}C_xC_y\right) + \left(-\lambda C_x^2 - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right)\right)$$

$T_n$ is approximately unbiased if the value of the constant is

$$\alpha = \left(\frac{\left(-\lambda C_x^2 - \lambda'\rho_{yx}C_xC_y + \frac{\lambda' C_x^2}{2}\right)}{\left(\lambda\rho_{yx}C_xC_y\right)}\right) \quad (6)$$

To obtain the error function of the estimator, we can rewrite equation (4) to have

$$T_n = \overline{Y}\left(\alpha\left(\varepsilon_0 + \varepsilon_1 - \varepsilon_1' - \varepsilon_1\varepsilon_1' + \varepsilon_0\varepsilon_1 - \varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2} - \varepsilon_0 - \varepsilon_2 + \varepsilon_1' + \varepsilon_2\varepsilon_1' + \varepsilon_0\varepsilon_1' - \frac{(\varepsilon_1')^2}{2}\right) + \left(\varepsilon_0 + \varepsilon_2 - \varepsilon_1' - \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right)$$

$$T_n = \overline{Y}\left(\alpha(\varepsilon_1 - \varepsilon_2 - \varepsilon_1\varepsilon_1' + \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1') + \left(\varepsilon_0 + \varepsilon_2 - \varepsilon_1' - \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right) \quad (7)$$

Squaring both sides of equation (7) and neglecting terms of ε's involving power greater than two, we have

$$\left(T_p - \overline{Y}\right)^2 = \overline{Y}^2\left(\alpha(\varepsilon_1 - \varepsilon_2 - \varepsilon_1\varepsilon_1' + \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1') + \left(\varepsilon_0 + \varepsilon_2 - \varepsilon_1' - \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right)^2$$

$$= \overline{Y}^2\left(\alpha^2(\varepsilon_1 - \varepsilon_2 - \varepsilon_1\varepsilon_1' + \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1')^2 + \left(\varepsilon_0 + \varepsilon_2 - \varepsilon_1' - \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)^2 + 2\alpha(\varepsilon_1 - \varepsilon_2 - \varepsilon_1\varepsilon_1' + \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1')\left(\varepsilon_0 + \varepsilon_2 - \varepsilon_1' - \varepsilon_2\varepsilon_1' - \varepsilon_0\varepsilon_1' + \frac{(\varepsilon_1')^2}{2}\right)\right)$$

$$= \overline{Y}^2\big(\alpha^2(\varepsilon_1^2 + \varepsilon_2^2 - 2\varepsilon_1\varepsilon_2) + (\varepsilon_0^2 + \varepsilon_2^2 + 2\varepsilon_0\varepsilon_2 + 2\varepsilon_0\varepsilon_1' - \varepsilon_2\varepsilon_1' + (\varepsilon_1')^2) + 2\alpha(-\varepsilon_2^2 + \varepsilon_0\varepsilon_1 - \varepsilon_0\varepsilon_2 + \varepsilon_1\varepsilon_2 - \varepsilon_1\varepsilon_1' + \varepsilon_2\varepsilon_1')\big) \quad (8)$$

Taking expectation on both sides on (8), we shall obtain the MSE of the estimator to the first degree of approximation as:

$$MSE(T_p) = \overline{Y}^2\Big(\alpha^2(\lambda C_x^2 + \lambda^* C_{x2}^2 + \lambda C_x^2 - 2\lambda C_x^2) + \left(\lambda C_y^2 + \lambda^* C_{y2}^2 + \lambda C_x^2 + 2\lambda\rho_{yx}C_xC_y + 2\lambda'\rho_{yx}C_xC_y - \lambda' C_x^2 + \lambda' C_x^2\right) - 2\alpha\left(\lambda C_x^2 - \lambda\rho_{yx}C_xC_y - \lambda\rho_{yx}C_xC_y - \lambda C_x^2 - \lambda' C_x^2 + \lambda' C_x^2\right)\Big)$$

$$MSE = \overline{Y}^2\left(\alpha^2(\lambda^* C_{x2}^2) + \left(\lambda C_y^2 + \lambda^* C_{y2}^2 + \lambda C_x^2 + 2\lambda\rho_{yx}C_xC_y + 2\lambda'\rho_{yx}C_xC_y\right) - 4\alpha\left(\lambda\rho_{yx}C_xC_y\right)\right) \quad (9)$$

The MSE obtained as equation (10) is minimized for

$$\alpha = \frac{\lambda\rho_{yx}C_xC_y}{\lambda^* C_{x2}^2}$$

Hence, the optimal value of α is

$$\alpha^{opt} = \frac{\lambda\rho_{yx}C_xC_y}{\lambda^* C_{x2}^2}$$

The optimal variance is

$$MSE(T_p)^{opt} = \overline{Y}^2\left(\lambda^* C_{x2}^2\left(\lambda C_y^2 + \lambda^* C_{y2}^2 + \lambda C_x^2 + 2\lambda\rho_{yx}C_xC_y + 2\lambda'\rho_{yx}C_xC_y\right) - 3\left(\lambda\rho_{yx}^2C_y^2C_x^2\right)\right) \quad (10)$$

**2.3. Description of Dataset used in the study**

Using auxiliary variables, this work examined five natural populations from three distinct sources to assess the effectiveness of the proposed modified ratio estimators and associated existing ratio estimators for population mean. To evaluate the performance of the estimators, we specifically examined two populations (populations 1 and 2) from [26], two populations (populations 3 and 4) from [27], and one population (population 5) from [22].

Data from [26] was presented as:

**Population 1:**

Y = Output for 80 factories in a region and X= Number of workers

N=80, n=20, $\overline{Y} = 51.8264$, $\overline{X} = 11.2646$, $\rho_{yx} = 0.9413$, $\rho_{yx2} = 0.8923$, $C_y = 0.3543$
$C_x = 0.7507$, $B_{1x}$=1.0500, $B_{2x}$= -0.0634, $M_d$=7.5750, $\lambda = 0.0375$, $\lambda^* = 0.05$, $\lambda' = 0.9625$

**Population 2:**

Y = Output for 80 factories in a region and X= Fixed Capital
N=80, n=20, $\overline{Y} = 51.8264$, $\overline{X} = 11.2646$, $\rho_{yx} = 0.9413$, $\rho_{yx2} = 0.8923$, $C_y = 0.3542$, $C_x = 0.9485$,
$C_{y2} = 1.1148$, $B_{1x}$=1.0500, $B_{2x}$= -0.0634, $M_d$=7.5750, $\lambda = 0.0375$, $\lambda^* = 0.05$, $\lambda' = 0.9625$

Data from [27] was presented as:

**Population 3:**

N=40, n=8, $\overline{Y} = 50.7858$, $\overline{X} = 2.3033$, $\rho_{yx} = 0.8006$, $\rho_{yx2}$ =0.7516, $C_y = 0.3295$, $C_x = 0.8406$, $C_{y2} =$
$0.1996$, $C_{x2} = 0.767$ $B_{1x}$=0.8799, $B_{2x}$= -0.4622, $M_d$=1.2500, $\lambda = 0.02$, $\lambda^* = 0.375$, $\lambda' = 0.1246$

Y = Output for 40 factories in a region and X = Number of workers

**Population 4:**

*Y* = Output for 40 factories in a region and X = Fixed Capital

N=40, n=8, $\overline{Y} = 50.7858$, $\overline{X} = 9.4543$, $\rho_{yx} = 0.8349$, $\rho_{yx2}$ =0.7859, $C_y = 0.3295$, $C_x = 0.6756$, $C_{y2} =$
$0.2095$, $C_{x2} = 0.6056$, $B_{1x}$=0.8799, $B_{2x}$= -0.4622, $M_d$=7.070, $\lambda = 0.02$, $\lambda^* = 0.375$, $\lambda' = 0.1246$

Data from [22] was presented as:

**Population 5**

Y= The number of agricultural labors and X= the number of cultivators
N=96, n=40, $\overline{Y} = 137.9271$, $\overline{X} = 144.872$, $\rho_{yx} = 0.773$, $\rho_{yx2}$ =0.724, $C_y = 0.3295$, $C_x = 0.8115$, $C_{y2} =$
$2.0838$, $C_{x2} = 0.9408$, $B_{1x}$=9.9541, $B_{2x}$= 101.1942, $M_d$=144.872, $\lambda = 0.0146$, $\lambda^* = 0.0062$, $\lambda' = 0.0038$

## 3. RESULTS AND DISCUSSIONS

The mean square error (MSE), bias, and constants for the proposed estimator and the several estimators that were considered in the study were assessed and then put through the Kruskal-Wallis and Shapiro-Wilk normality tests. The normality of the constants, bias, and mean square error (MSE) was assessed in the study using the Shapiro-Wilk normality test. It was evident from the test findings that none of the three data sets was normally distributed since the *p*-values were less than the significant threshold of 0.05. As a result, a non-parametric technique like the Kruskal-Wallis Test was used to evaluate the variability in the performance of the estimators for the mean value of constants, mean value of bias, and mean value of MSE.

Table 1. Summary result of Kruskal-Wallis Test for the measure of the Constant, Bias and MSE of the Estimators

| S/No. | Performance Indices | Kruskal-Wallis chi-squared | Degree of Freedom (df) | *p*-value |
|---|---|---|---|---|
| 1 | Constant | 24.114 | 10 | 0.0073 |
| 2 | Bias | 24.586 | 10 | 0.0062 |
| 3 | MSE | 11.316 | 10 | 0.3334 |

Source: Authors Analysis

The findings displayed in Figure 1 demonstrate that the mean values of the constants among the different estimators taken into consideration in the study vary significantly. The SA estimator then validated the JK estimator's lowest mean value for the constant. The proposed estimator yielded the constant's third-least mean value. On the other hand, the IET estimator was found to record the greatest mean value for the constant. The results indicate that the JK and SA estimators outperformed other existing estimators considered in the study when estimating the population means using an auxiliary variable that considered the constant's mean value.  Though the proposed estimator was not absolute the better performer in this regard, it still showcased competitive performance. Hence, the findings indicates that the proposed estimator offers an accurate estimations of constants, which are crucial for population mean estimation.

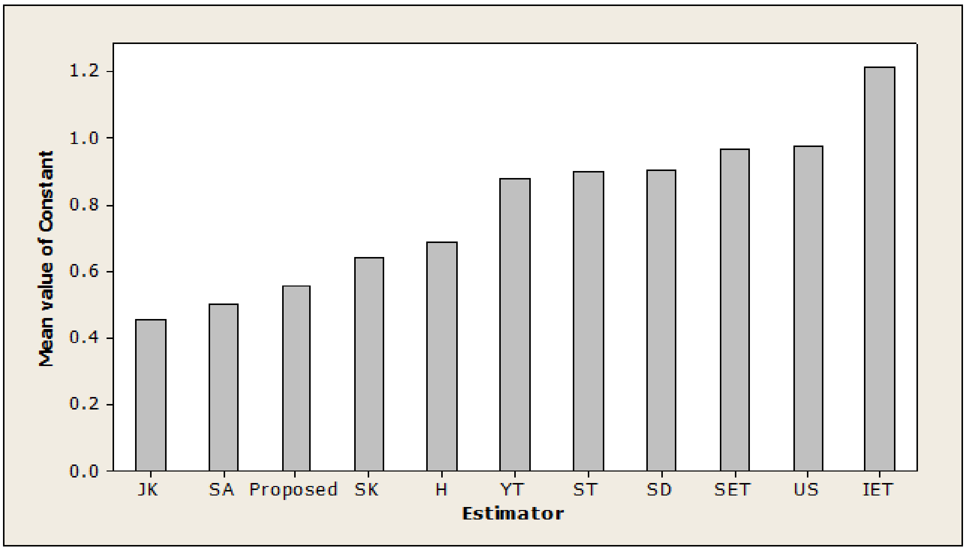

Figure 1: Mean values of the constant of the Estimators

The outcome displayed in Figure 2 indicates that there is a significant variation in the estimators' mean values of bias when it comes to their performance in the study. The US estimator has the largest mean value of bias, while the proposed estimator has the lowest mean value, followed by the H estimator. Due to their lower degrees of bias when compared to other current estimators taken into consideration for the study, this finding implies that the proposed estimator and the H estimator produce more accurate and exact estimations of the population mean utilizing an auxiliary variable. This outcome indicates that the proposed estimator provides a more accurate and precise estimations of the population mean, with minimized bias compared to other estimators. This implies that the performance of the proposed estimator in minimizing bias surpassed that of the existing estimators like the US estimator, which showed the largest mean value of bias.

Based on the results displayed in Figure 3, the proposed estimator has the lowest mean MSE value, followed by the JK estimator, and the H estimator has the highest mean MSE value. With lower values of MSE than other current estimators taken into consideration in the study, this finding implies that the proposed estimator recorded a better estimate of the population mean by utilizing an auxiliary variable. Hence, the proposed estimator shows how well it can estimate population means with accuracy. This indicates that the proposed estimator performed competitively, establishing it as a good choice for precise mean estimate even in the face of other estimators with lower MSE values.

With $p$-values of 0.0073 and 0.0062 (both of which are less than the crucial threshold of 0.05), the data displayed in Table 1 demonstrate that there is a statistically significant difference in the distribution of mean values for the constants and bias. Nonetheless, no noteworthy distinction was noted in the average MSE values of the estimators. Dunn multiple comparison tests must be run in order to pinpoint the precise pair of estimators that are in charge of the observed variation in the mean values of constants and bias.

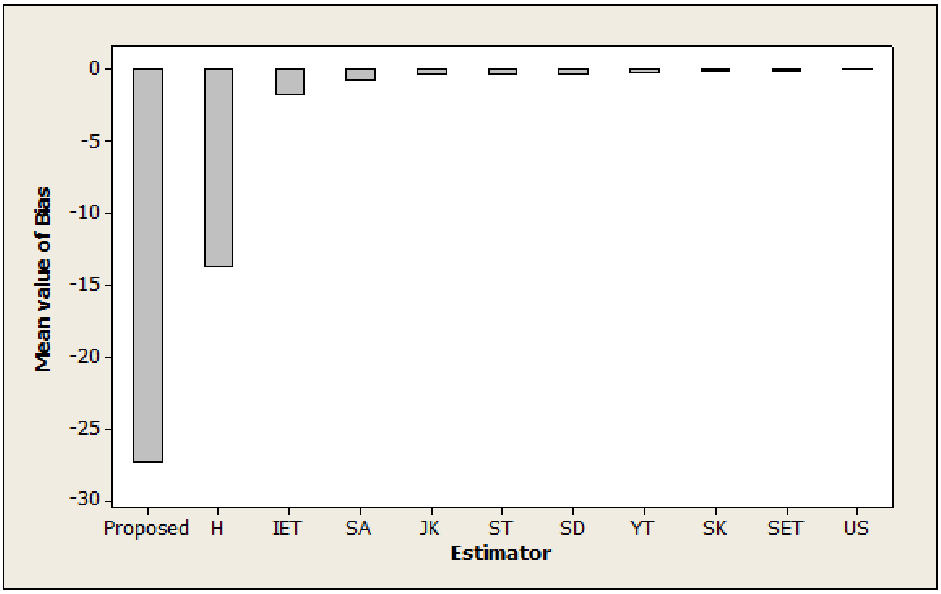


Figure 2. Mean values of the Bias of the Estimators

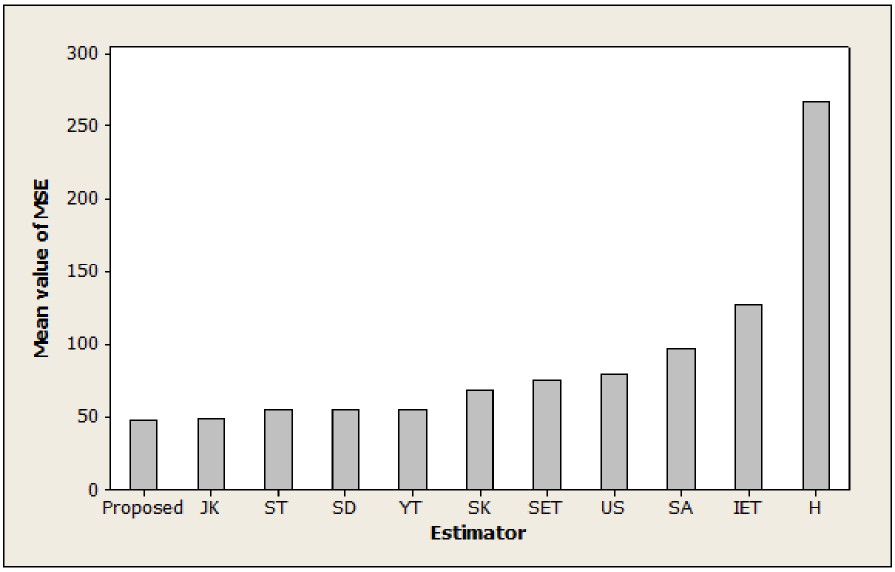


Figure 3. mean values of the MSE of the Estimators

The Dunn Kruskal-Wallis multiple comparison test with the Bonferroni correction method was used to analyze the distribution of the measure of the constants. The following pairs of variables showed significant differences: proposed-ST and SA-ST ($p - value = 0.0314$), proposed-US and SA-US ($p - value = 0.0102$), SD-proposed and SD-SA ($p - value = 0.0232$), SA-SET and proposed-SET ($p - value = 0.0097$ and $0.0060$, respectively), US-SK and SET-SK ($p - value = 0.0364$ and $0.0346$, respectively), SD-JK and US-JK ($p - value = 0.0198$ and $0.0053$, respectively), ST-JK and SET-JK ($p - value = 0.0270$ and $0.005$, respectively), SA-IET and proposed-IET ($p$-value=0.0461 and 0.0314, respectively), and JK-IET ($p$-value= 0.0270). Hence, the Dunn Kruskal-Wallis test with Bonferroni method indicates that there are statistically significant differences among the pairs of estimators as stated above.

Similarly, the Dunn Kruskal-Wallis multiple comparison test with the Bonferroni method was used to analyze the distribution of the measure of Bias. The test revealed significant differences between several pairs of estimators. Specifically, there were significant differences between SD-proposed ($p - value = 0.022$), proposed-US ($p - value = 0.0109$), proposed-ST ($p - value = 0.0232$), proposed-SET ($p$-value=0.0129), proposed-YT ($p$-value=0.0285), and proposed-SK ($p$-value=0.0441). Additionally, there were significant differences between SD-H ($p - value = 0.0031$), US-H ($p - value = 0.0013$), ST-H ($p - value = 0.0033$), SET-H ($p - value = 0.0016$), YT-H ($p - value = 0.0042$), and SK-H ($p - value = 0.0073$). The test also revealed a significant difference between H-JK ($p - value = 0.033$), US-IET ($p - value = 0.0364$), and SET-IET ($p - value = 0.042$). These findings suggest that there are significant differences between the Bias values of the various pairs of estimators listed above. This implies that the choice of the estimator can have a significant impact on the Bias value. Therefore, researchers and practitioners should carefully consider their choice of the estimator when analyzing Bias.

In addition, Dunn Kruskal-Wallis multiple comparison test with the Bonferroni correction method was used to analyze the distribution of the measure of the MSE. The outcome revealed that there is significant difference between the following pair of variables SD-proposed($p$-value=0.022), proposed-US ($p - value = 0.0109$), proposed-ST ($p - value = 0.0232$), proposed-SET($p - value = 0.0129$), proposed-YT($p - value = 0.0285$), proposed-SK($p - value = 0.0441$), SD-H($p - values = 0.0031$), US-H($p - value = 0.0013$), ST-H($p - value = 0.0033$), SET-H($p - value = 0.0016$), YT-H($p$-value= 0.0042), SK-H($p$-value=0.0073), H-JK($p$-value=0.033), US-IET ($p$-value=0.0364), and SET-IET ($p - value = 0.042$). The significant differences in MSE between the proposed estimator and other estimators considered in the study, as found by the Dunn Kruskal-Wallis multiple comparison test, implies that the proposed estimator outperformed other existing estimators in terms of accuracy and precision in estimating the population mean. These outcomes provides useful novel information on how well the proposed estimator performed in comparison to other estimators.

## 4. CONCLUSION

This study proposed a ratio-type estimator that is designed for estimating population means in cases of non-response, by employing twofold sampling and an auxiliary variable. The main findings of the research contribute significantly to the understanding and application of ration-type estimator, particularly in situations involving double sampling and non-response. The study found significant variations in the average constant, bias, and mean square error values across the various estimators examined. The proposed estimator, along with the $H$ estimator, were found to express more exact and accurate estimates of the population mean utilizing an auxiliary variable compared to other existing estimators considered in the study. The proposed estimator exhibits lower mean square error (MSE) values compared to other estimators, indicating its superior performance in providing better estimates of the population mean using an auxiliary variable. Also, significant differences was observed in the distribution of constant, bias, and MSE values among the estimators, as revealed by the Dunn Kruskal-Wallis multiple comparison test, this implies the statistical significance of the findings. In addition, the study provides valuable insights for researchers and practitioners, guiding them in selecting appropriate estimators for specific scenarios. Moreover, the findings suggest that the proposed estimator, with its reduced bias and MSE values, can lead to more dependable population mean estimations, especially in situations where precise and accurate estimates are crucial. By introducing the unique expressions for constants, bias, and mean square error specific to the proposed estimator, the research advances the theoretical framework underlying population estimation methodologies. The demonstrated superiority of the proposed estimator in terms of reduced bias and mean square error suggests its practical utility in real-world scenarios where accurate population estimates are crucial for decision-making processes. The research contributes to the generation of reliable demographic data, which forms the basis for evidence-based policy formulation and implementation. The extent to which the findings of the study can be extrapolated or transferred to other contexts, populations, or time periods should be critical considered. This is because while the study may yield valuable insights within its specific context, caution should be exercised when applying the findings to different situations or populations. It is expected that factors such as cultural differences, policy frameworks, or technological advancements may influence the transferability of the research findings.

## REFERENCES

[1] A. Kumar, & R.S. Bhougal, "A modified ratio-product exponential estimator for population mean in the presence of non-response", Journal of Statistics and Management Systems, 14(2), 261-278, 2011.

[2] R.P.B. Singh, B. Sinha, & R. Kumar, "Estimation of population mean in the presence of non-response: a comparative study", Journal of Applied Statistical Science, 16(3), 265-274, 2008.

[3] J.N.K. Rao, "Estimation of population mean and variance in the presence of nonresponse", Journal of the American Statistical Association, 81(393), 1010-1016, 1986.

[4] R.P. Singh, & R. Kumar, "Estimation of population mean in the presence of non-response: A comparative study", Journal of Probability and Statistics, 1-10, 2009.

[5] O.A. Sodipo, & A.A. Obisesan, "A generic class of estimators for population mean in the presence of non-response", Statistics in Transition, 9(3), 471-484, 2007.

[6] R. Kumar, "Estimation of population mean in presence of non-response: An improvement", Journal of Applied Statistics, 40(5), 1007-1016, 2013.

[7] B. Sharma, & H.P. Singh, "Estimation of population mean using two auxiliary variables", Communications in Statistics-Simulation and Computation, 44(5), 1364-1384, 2015.

[8] M.Z. Asghar, M.A. Rao, & M. Khalid, "Generalized class of exponential-type estimators for the population variance using known value(s) of population mean of auxiliary variable(s)", Journal of Applied Statistics, 41(11), 2392-2407, 2014. https://doi.org/10.1080/02664763.2014.907559.

[9] M. Kaur, & L.K. Grover, "Exponential ratio type estimators of population mean in the presence of missing data for both study and auxiliary variable", Journal of Probability and Statistical Science, 11(2), 115-131,

2013. https://doi.org/10.1016/j.jkss.2013.05.001
[10] L.K. Grover, & M. Kaur, "Exponential ratio type estimators of population mean in the presence of missing data", Metron, 72(1), 53-68, 2014. https://doi.org/10.1007/s40300-014-0031-6
[11] D. Kumar, & M. Viswanathaiah, "Improved exponential ratio-type estimators for population mean in double sampling", Journal of Applied Statistics, 41(9), 1959-1973, 2014. doi: 10.1080/02664763.2014.890232
[12] D. Kumar, H.K. Singh, & G. Garg, "Estimation of population mean of response variable in presence of non-response and measurement errors using auxiliary information", International Journal of Applied and Computational Mathematics, 1(3), 197-209, 2015. doi: 10.1007/s40819-015-0046-2
[13] Liu, H., & Singh, H. P. "New exponential-type estimators using auxiliary information in sample surveys with non-response", Journal of Applied Statistics, 44(11), 1961-1979, 2017.
[14] M.M. Rahman, M.A. Islam, & M.A.R. Ahad, M. A. R. "A new weighted estimator using auxiliary information in sample surveys with non-response", Journal of Applied Statistics, 46(10), 1916-1935, 2019.
[15] P. Kaur, & H.P. Singh, "A new estimator of population mean in sample surveys with non-response", Statistical Papers, 61(3), 1049-1065, 2020.
[16] N.S. Sisodia, & S.N. Dwivedi, "Estimation of population mean in the presence of measurement errors in auxiliary variable", Sankhya: The Indian Journal of Statistics, Series B, 43(2), 142-149, 1981.
[17] L.N. Upadhyaya, & H.P. Singh, "On some ratio-type estimators in sample surveys", Metrika, 49(2), 165-175, 1999.
[18] H.P. Singh, & R. Tailor,"Use of auxiliary information in estimating population mean using two-phase sampling", Journal of Applied Statistical Science, 12(1), 55-67, 2003.
[19] H.P. Singh, R. K. Grewal, & P. Kumar, "Improvement in estimating population mean using known correlation coefficient between an auxiliary variable and a study variable", Journal of Applied Statistical Science, 13(2), 155-167, 2004.
[20] Z. Yan, & W. Tain, "A new class of estimators for population mean in simple random sampling", Journal of Applied Statistics, 37(10), 1651-1661, 2010.
[21] J. Subramani, & M.A. Krumarpandiyan, "Improved ratio estimators using auxiliary information in simple random sampling", Journal of Applied Statistical Science, 21(1), 123-132, 2013.
[22] R. Hazara, "Combination of Two Ratio Type Estimators for Estimating Population Mean Using Auxiliary Variable with Double Sampling in Presence of Non- Response", Bulletin of Mathematics and Statistics Research, 3(2): 33-41, 2015.
[23] A. Jerajuddin, & R. Kishun, "On some improved ratio-type estimators in survey sampling", Journal of Statistics Applications & Probability, 5(1), 1-16, 2016.
[24] M. Ijaz, A.U. Khan, S.M. Asim, Y. Hayat, S.H. Shah, M. Farooq, A.K. Neelam, and S. Zubair, "A New Ratio Type Estimator to Estimate the Population Mean using Auxiliary Information", Advances and Applications in Statistics, 63(1): 97-108, 2020.
[25] M.M. Suleiman & A.A. Adewara, "A new ratio estimator for population mean using coefficient of variation of an auxiliary variable", Heliyon, 7(7), e07647, 2021. https://doi.org/10.1016/j.heliyon.2021.e07647
[26] M.N. Murthy, "Sampling Theory and Methods", Statistical Publishing Society, Calcutta, 1967.
[27] P. Mukhopadhyay, "Theory and methods of survey sampling", PHI Learning, 2nd edition, New Delhi, 2009.