# A Refined Population Mean Estimator Using Median and Skewness: Applications to Breast Cancer and Brain Tumor Data

N. Venkata Lakshmi[1], Faizan Danish[1], Mustafa Ibrahim Ahmed Araibi[2,*], I. Elbatal[3], Ehab M. Almetwally[3], Ahmed M. Gemeay[4], Sonali Kedar Powar[5] and Aafaq A. Rather[6]

[1]Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology-Andhra Pradesh (VIT- AP) University, Inavolu, Beside AP Secretariat, Amaravati AP-522237, India

[2]Department of Business Admiration, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

[3]Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

[4]Department of Mathematics, Faculty of Science, Tanta University, Tanta 31527, Egypt

[5]Department of Computer Science, Faculty of Science and Technology, Vishwakarma University, Pune, India

[6]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India

**Abstract:** Estimators are essential to sampling theory because they allow researchers and statisticians to calculate estimates of population parameters from observed data. In every survey activity, the experimenter aims to use methods that will improve the precision of population parameter estimations throughout both the design and estimation phases. When auxiliary data is used in the estimating, design, or both processes, these estimated precisions can be attained. By linearly merging the central value of the data under consideration with the skewness coefficient provided by Karl Pearson, this study created a new, improved predictor for calculating the average of a population. Estimators are crucial to sampling theory because of their capacity to produce estimates of population parameters from observed data.

In this work, a novel modified ratio-type estimator was constructed by linearly merging Karl-Pearson's coefficient of skewness with the median value. Simple random sampling (SRS) was the technique employed in this present study. We conduct a numerical analysis from the standpoint of real estate. Additionally, we do some real data analysis on two distinct cancers: the brain tumor dataset and the breast cancer dataset. The results of the simulation study, real data application in the medical field, and numerical investigation show that the suggested estimator achieves lower error when the median value and Karl Pearson's coefficient of skewness are combined. Furthermore, compared to the other estimators under consideration, the one proposed in this study achieves better precision.

**Keywords:** Ratio-estimator, Simple Random Sampling, Mean Square Error.

## INTRODUCTION

Sampling theory is the foundation for concluding about populations from sample data in statistical analysis. The application of sampling techniques can result in speedier outcomes, greater scope, greater precision, and cost savings. The process you use to choose your samples is known as sampling design. When choosing your survey sample, you can use any of the many different types of sampling designs as a reference. We suggested an estimator in this study's simple random sampling approach. Every member of the population has an equal probability of being selected using a straightforward random sampling technique. Estimators are essential tools that enable practitioners and researchers to draw insightful inferences from sparse data. A ratio estimator is one specific type of estimator in survey sampling that was introduced in the first study [18] to show how having auxiliary information might improve efficiency when computing the population mean. The estimation method described in [18] is more beneficial when the correlation between the primary variable and the associated variable is greater and non-negative.

Using a known variation coefficient of the associated variable, [1] proposed a modified ratio estimator. They develop their ratio-type estimators by taking into consideration the kurtosis coefficient and the coefficient of variation [3]. The ratio estimators were applied to a study on apple productivity by [4]. They came to the conclusion that the classical ratio estimator is the most successful of all the estimators used in the comparison analysis. An estimate created by [5] for ratios was given using a known correlation value.

The problem of calculating the population mean of the study variate was addressed by a ratio estimator developed by [8] using the coefficient of variation of an auxiliary character. A modified ratio estimate created by

*Address correspondence to this author at the Department of Business Admiration, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia; E-mail: miarbi@imamu.edu.sa

E-ISSN: 1929-6029/25

[9] is based on the correlation coefficient and takes into account the estimators of [3]. An estimate developed by [2] provides a class of ratio-type estimators of the population means in finite population sample surveys using Simple Random Sampling (without replacement) when data on an auxiliary variate positively associated with the research variable is available. Auxiliary data is used in SRS design by a family of estimators created by [10]. [11] proposed some ratio-type estimators for the population mean based on the skewness coefficient of the connected variable. Inspired by [7, 12] provided some effective sampling techniques to estimate the population mean of the research variable in a finite population by using the kurtosis coefficient of the associated variable. [13] Introduced an alternative ratio estimator by constructing a linear mixture of the estimators in [5] and [6]. The estimators developed by [14] combine data on sample size, kurtosis coefficients, and non-traditional metrics with quartile deviation that is not impacted by outliers to provide an estimator with higher precision. A generalized family of Exponential Factor-Type Estimators for the Population Distribution Function was developed by [15] using dual estimator auxiliaries. [16] used the Simple Random Sampling technique to offer a composite class of ratio estimators for the population mean of a study variable. In [20], a log-type of ratio estimators is proposed for estimating a population's finite mean. Three alternative estimators are tested using the SRS technique in a simulation in [17]. [23] used their estimators to support their theoretical results in four numerical examples from agricultural, biomedical, and power engineering, extending survey sample theory by developing new estimators with two auxiliary variables. [24] Recommended an estimator and shown its use in economics, medicine, and demographic studies.

In this study, we employ the median and Karl Pearson's coefficient of skewness to improve the accuracy of the proposed estimator, particularly when handling outliers or skewed data. In situations where the data was not symmetrically distributed, the coefficient of skewness reduced bias by accounting for distributional asymmetry. Compared to traditional estimation methods which uses skewness and median in earlier studies, the estimator advised in this article became more robust to outliers by using the median, which is less prone to extreme values than the mean. This produced a more reliable estimate of central tendency. We demonstrated this by applying estimators to real data sets on brain tumor and breast cancer [21, 22]. Both simulation and numerical research methods attest to the fact that this combination increases the estimator's effectiveness and makes it more appropriate for the use of the suggested estimator in real-world datasets.

## Notations and Terminology

Consider a sample of size n drawn from the given N-unit population. Let Y be the value of the variable being investigated, and X be the associated variable under discussion.

The formulas listed below provide the essential parameters and data for this inquiry.

$\bar{y} = \frac{1}{n}\sum_{i=}^{n} y_i$ , the sample mean of the auxiliary variable (X).

$\bar{x} = \frac{1}{n}\sum_{i=}^{n} x_i$, the sample mean of the study variable(Y).

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, sample variance of X.

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{x}(x_i - \bar{x})^2$, sample variance of Y.

$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$, sample covariance of X and Y

Here $y_i$ and $x_i$ represent the $i^{th}$(i=1, 2, 3, …, n) units from the samples of the study and auxiliary variables, respectively

$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$, population average of Y.

$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, population average of X.

$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$, The variance of Y based on the population.

$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$, variance of X based on the population.

$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})$ Covariance between X and Y based on the population.

Here $Y_i$ and $X_i$ represent the $i^{th}$(i=1, 2, 3, …, N) units from the populations of research and related variables, respectively

And also

$C_y^2 = \frac{S_y^2}{\bar{Y}^2}$, $C_x^2 = \frac{S_x^2}{\bar{X}^2}$, and $\rho_{xy} = \frac{S_{xy}}{S_x S_y}$ where $S_x = sqrt(S_x^2)$, $S_y = sqrt(S_y^2)$, $C_x = sqrt(C_x^2)$ and $C_y = sqrt\left(S_y^2\right)$

$\beta_{2(x)} = \frac{n(n-1)}{(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$;

$\beta_{1(x)} = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)^3$, $s_x$ is the sample standard deviation.

the assumptions for the error terms are [19]

$\bar{y} = (1+e_0)\bar{Y}$ and $\bar{x} = (1+e_1)\bar{X}$

For notation simplicity, let us take

$(Ee_0) = E(e_1) = 0$; $E(e_0^2) = \frac{1-f}{n}C_y^2 = \Delta_0$; $E(e_0e_1) = \frac{1-f}{n}\rho_{xy}C_xC_y = \Delta_1$ ;

$E(e_1^2) = \frac{1-f}{n}C_x^2 = \Delta_{01}$; $f = \frac{n}{N}$ , denotes the finite population correction factor.

Where n is the sample size and N is the population size.

### Existing Estimators

1. The combined ratio estimator by [18] is defined as

$$\bar{y}_1 = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{1}$$

The MSE of (1) is

$$MSE(\bar{y}_1) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + C_x^2 - 2\rho_{xy}C_xC_y\right] \tag{2}$$

Also, the bias of (1) is

$$B(\bar{y}_1) = \left(\frac{1-f}{n}\right)\bar{Y}\left[C_x^2 - \rho_{xy}C_xC_y\right] \tag{3}$$

2. When the variation coefficient is available, a modified ratio estimator [1] is suggested for $\bar{y}\ as$

$$\bar{y}_2 = \bar{y}\frac{\bar{X}+C_x}{\bar{x}+C_x} \tag{4}$$

The MSE of (4) is

$$MSE(\bar{y}_2) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_1}^2C_x^2 - 2\theta_1\rho_{xy}C_xC_y\right] \tag{5}$$

also, the bias of (4) is

$$B(\bar{y}_2) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_1}^2C_x^2 - \theta_1\rho_{xy}C_xC_y\right] \tag{6}$$

here $\theta_1 = \frac{\bar{X}}{\bar{X}+C_x}$;

3. A modified ratio estimator was proposed by [5] as

$$\bar{y}_3 = \bar{y}\frac{\bar{X}+\rho}{\bar{x}+\rho} \tag{7}$$

The MSE of (7) is

$$MSE(\bar{y}_3) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_2}^2C_x^2 - 2\theta_2\rho_{xy}C_xC_y\right] \tag{8}$$

Also, the bias of (7) is

$$B(\bar{y}_3) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_2}^2C_x^2 - \theta_2\rho_{xy}C_xC_y\right] \tag{9}$$

where $\theta_2 = \frac{\bar{X}}{\bar{X}+\rho}$

here $\rho$ is the correlation coefficient.

4. A ratio-type estimator was proposed by [3] is defined as

$$\bar{y}_4 = \bar{y}\frac{\bar{X}C_x+\beta_{2(x)}}{\bar{x}C_x+\beta_{2(x)}} \tag{10}$$

The MSE of (10) is

$$MSE(\bar{y}_4) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_3}^2C_x^2 - 2\theta_3\rho_{xy}C_xC_y\right] \tag{11}$$

Also, the bias of (10) is

$$B(\bar{y}_4) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_3}^2C_x^2 - \theta_3\rho_{xy}C_xC_y\right] \tag{12}$$

here $\theta_3 = \frac{\bar{X}C_x}{\bar{X}C_x+\beta_{2(x)}}$;

5. By switching the locations of the coefficient of variation and the coefficient of kurtosis, another ratio estimator proposed by [3] as

$$\bar{y}_5 = \bar{y}\frac{\bar{X}\beta_{2(x)}+C_x}{\bar{x}\beta_{2(x)}+C_x} \tag{13}$$

The MSE of (13) is

$$MSE(\bar{y}_5) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_4}^2C_x^2 - 2\theta_4\rho_{xy}C_xC_y\right] \tag{14}$$

Also, the bias of (13) is

$$B(\bar{y}_5) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_4}^2C_x^2 - \theta_4\rho_{xy}C_xC_y\right] \tag{15}$$

where $\theta_4 = \frac{\bar{X}\beta_{2(x)}}{\bar{X}\beta_{2(x)}+C_x}$;

6. A ratio-type estimator was suggested by [11] as

$$\bar{y}_6 = \bar{y}\frac{\bar{X}+\beta_{1(x)}}{\bar{x}+\beta_{1(x)}} \tag{16}$$

The MSE of (16) is

$$MSE(\bar{y}_6) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_5}^2C_x^2 - 2\theta_5\rho_{xy}C_xC_y\right] \tag{17}$$

Also, the bias of (16) is

$$\mathrm{B}(\bar{y}_6) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_5}^2 C_x^2 - \theta_5 \rho_{xy} C_x C_y\right] \tag{18}$$

where $\theta_5 = \frac{\bar{X}}{\bar{X}+\beta_{1(x)}}$;

7. A ratio-type estimator was suggested by [11] as

$$\bar{y}_7 = \bar{y}\frac{\bar{X}\beta_{1(x)}+\beta_{2(x)}}{\bar{x}\beta_{1(x)}+\beta_{2(x)}} \tag{19}$$

The MSE of (19) is

$$\mathrm{MSE}(\bar{y}_7) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_6}^2 C_x^2 - 2\theta_6 \rho_{xy} C_x C_y\right] \tag{20}$$

Also, the bias of (19) is

$$\mathrm{B}(\bar{y}_7) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_6}^2 C_x^2 - \theta_6 \rho_{xy} C_x C_y\right] \tag{21}$$

where $\theta_6 = \frac{\bar{X}\beta_{1(x)}}{\bar{X}\beta_{1(x)}+\beta_{2(x)}}$;

8. Based on t6he coefficients of variation and skewness [11] provided the following estimator

$$\bar{y}_8 = \bar{y}\frac{\bar{X}C_x+\beta_{1(x)}}{\bar{x}C_x+\beta_{1(x)}} \tag{22}$$

The MSE of (22) is

$$MSE(\bar{y}_8) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_7}^2 C_x^2 - 2\theta_7 \rho_{xy} C_x C_y\right] \tag{23}$$

Also, the bias of (22) is

$$B(\bar{y}_8) = \left(\frac{1-f}{n}\right)\bar{Y}\left[{\theta_7}^2 C_x^2 - \theta_7 \rho_{xy} C_x C_y\right] \tag{24}$$

Where $\theta_7 = \frac{\bar{X}C_x}{\bar{X}C_x+\beta_{1(x)}}$;

## Proposed Estimator

In this paper, we suggest an estimator that combines the median of the auxiliary variable and Karl Pearson's coefficient of skewness in a linear fashion. Since the median more accurately depicts the middle of a data collection than the mean, particularly in cases when the data is skewed or contains outliers, the median is a crucial indicator of central tendency, and it is less impacted by extreme numbers at either end of the data set. Karl Pearson's coefficient of skewness measures the asymmetry of a probability distribution and helps you understand the strength and direction of skewness in a sample distribution. The value of the coefficient of skewness can range from -3 to +3 and can be defined as

Positive value: The distribution is skewed to the right.

Negative value: The distribution is skewed to the left.

Value of 0: The distribution is symmetric.

The suggested estimator is defined as

$$\bar{y}_p = \bar{y}\frac{\bar{x}S_k+M_d}{\bar{X}S_k+M_d} \tag{25}$$

where $s_k = \frac{3(\bar{X}-M_d)}{S_x}$; $\bar{X}$ is the mean of the auxiliary variate;

$S_x$ is the population standard deviation of the auxiliary variable, and $M_d$ is the median of the auxiliary variable. When a data collection is organized in either ascending or descending order, the median is the midway value. Whether the data set has an odd or even number of values determines the formula for calculating the median.

If n=even, then $M_d$ = value at the position $\frac{n+1}{2}$

If n=odd, then $M_d = \frac{value\ at\ position\left(\frac{n}{2}\right)+\ value\ at\ position\left(\frac{n}{2}+1\right)}{2}$

The suggested ratio estimate's MSE is provided by

$$MSE(\bar{y}_p) = E(\bar{y}_p - \bar{Y})^2$$

$$MSE(\bar{y}_p) = E(\left[\bar{y}\frac{\bar{x}S_k+M_d}{\bar{X}S_k+M_d}\right]^\gamma - \bar{Y})^2$$

Now, consider error terms in place of $\bar{x}$ and $\bar{y}$, where $\bar{x} = (1+e_1)\bar{X}$ and $\bar{y} = (1+e_0)\bar{Y}$

$$MSE(\bar{y}_p) = E\left\{(1+e_0)\bar{Y}\left[\frac{\bar{x}S_k+M_d}{(1+e_1)\bar{X}S_k+M_d}\right]^\gamma - \bar{Y}\right\}^2$$

$$= E\left\{(1+e_0)\bar{Y}\left[\frac{\bar{X}S_k+\bar{X}S_k e_1++M_d}{\bar{X}S_k+M_d}\right]^\gamma - \bar{Y}\right\}^2$$

$$= E[(1+e_0)\,\bar{Y}(\bar{X}S_k+M_d)(\bar{X}S_k+M_d)^{-1}\,(1+e_1\theta_p)^\gamma - \bar{Y}]^2\ ,$$

$$= E[(1+e_0)\,\bar{Y}\,(1+e_1\theta_p)^\gamma - \bar{Y}]^2$$

where $\theta_p = \frac{\bar{X}S_k}{\bar{X}S_k+M_d}$

using the Binomial expansion of the terms $(1+e_1\theta_p)^\gamma$ we have

$$= E[(1+e_0)\,\bar{Y}(1+\gamma e_1\theta_p + \frac{\gamma(\gamma-1)}{2}{e_1}^2\theta_p^2) - \bar{Y}]^2$$

$$=E[\,\bar{Y}(1+\gamma e_1\theta_p + \frac{\gamma(\gamma-1)}{2}{e_1}^2\theta_p^2 + e_0 + \gamma e_0 e_1\theta_p) - \bar{Y}]^2$$

By neglecting the higher-order terms and cancelling the like-terms, we get

$MSE(\bar{y}_p) = \bar{Y}^2\left[\gamma^2 E({e_1}^2)\theta_p^2 + E({e_0}^2) + 2\gamma(e_0 e_1)\theta_p\right]$ (26)

$MSE(\bar{y}_p) = \bar{Y}^2\left[\gamma^2 \Delta_1 \theta_p^2 + \Delta_0 + 2\gamma\Delta_{01}\theta_p\right]$

To optimize the value of $\gamma$, Differentiate (26) concerning '$\gamma$' and equate it to zero, then we get

$$\frac{\partial\left[MSE(\bar{y}_p)\right]}{\partial\,\gamma} = 0$$

$$\gamma = \frac{-e_0 e_1}{{e_1}^2\theta_p}$$

Also, the bias of the proposed ratio estimate can be defined as

$B(\bar{y}_p) = E(\bar{y}_p) - \bar{Y}$

$= E[\bar{Y}(1 + \gamma e_1\theta_p + \frac{\gamma(\gamma-1)}{2}{e_1}^2\theta_p^2 + e_0 + \gamma e_0 e_1\theta_p] - \bar{Y}$

$$B(\bar{y}_p) = \bar{Y}\left[\frac{\gamma(\gamma-1)}{2}E({e_1}^2)\theta_p^2 + \gamma E(e_0 e_1)\theta_p\right]$$

$B(\bar{y}_p) = \bar{Y}\left[\frac{\gamma(\gamma-1)}{2}\Delta_1\theta_p^2 + \gamma\Delta_{01}\theta_p\right]$ (27)

where $\theta_p = \frac{\bar{X}S_k}{\bar{X}S_k + M_d}$, where $s_k$ is Karl Pearson's coefficient of skewness of an auxiliary variate.

### Theoretical Comparisons

The MSE of the suggested ratio estimator is contrasted with the MSE of the other estimators discussed in this article in this section: Using the suggested ratio-type estimator to compare (2), that is

$MSE(\bar{y}_p) - MSE(\bar{y}_1) < 0$ (28)

By solving the above equation, we get

$\rho_{xy} > \frac{(\theta_p+1)\,C_x}{2C_y}$ (29)

On comparing $MSE(\bar{y}_i)$, (i = 2,3, …, 8.) with $MSE(\bar{y}_p)$

The MSE of other ratio-type estimators under this article is given by

$MSE(\bar{y}_i) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_i}^2 C_x^2 - 2\theta_i\rho_{xy}C_x C_y\right], i =$ 2,3, ...,8.

$MSE(\bar{y}_p) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left[C_y^2 + {\theta_p}^2 C_x^2 - 2\theta_p\rho_{xy}C_x C_y\right]$

$MSE(\bar{y}_p) - MSE(\bar{y}_i) < 0, i = 2,3, \dots, 8.$ (30)

By solving the above equation (14) we get

$(\theta_p + \theta_i)C_x - 2\rho_{xy}C_y < 0$, i = 2,3, … ,8.

$2\rho_{xy}C_y > (\theta_p + \theta_i)C_x, i = 2,3, \dots, 8.$

$\rho_{xy} < \frac{(\theta_p+\theta_i)\,C_x}{2C_y}, i = 1,2, \dots, 8.$

$\rho_{xy} > \delta\,\frac{C_x}{C_y}$ (31)

where $\delta = \frac{(\theta_p+\theta_i)}{2}, i = 2,3, \dots, 8.$

when it met conditions (29) and (31), the suggested estimator $\bar{y}_p$ It is more efficient than other estimators $\bar{y}_i$, (i = 1,2, … … ,8).

**Table 1: Parameters and Constants of the Given Population**

| Parameters notation | Description of the parameter | Constants of the population |
|---|---|---|
| N | Size of the population | 50 |
| n | Size of the sample | 8 |
| $\bar{X}$ | Mean of the auxiliary variable | 878.16 |
| $\bar{Y}$ | Mean of the study variable | 555.43 |
| $\rho$ | Population correlation coefficient between the auxiliary and study variable | 0.8038 |
| $s_x$ | The standard deviation of the auxiliary variable of the population | 1073.776 |
| $M_d$ | The median of the auxiliary variable of the population | 452.517 |
| $\beta_1$ | Coefficient of skewness of the auxiliary variable of the population | 1.609783 |
| $\beta_2$ | Coefficient of kurtosis of the auxiliary variable of the population | 1.921544 |
| $s_k$ | Karl Pearson's coefficient of skewness of the auxiliary variable | 1.189194 |
| $C_x$ | Coefficient of variation of auxiliary variable | 1.23515 |
| $C_y$ | Coefficient of variation of the study variable | 1.05290 |

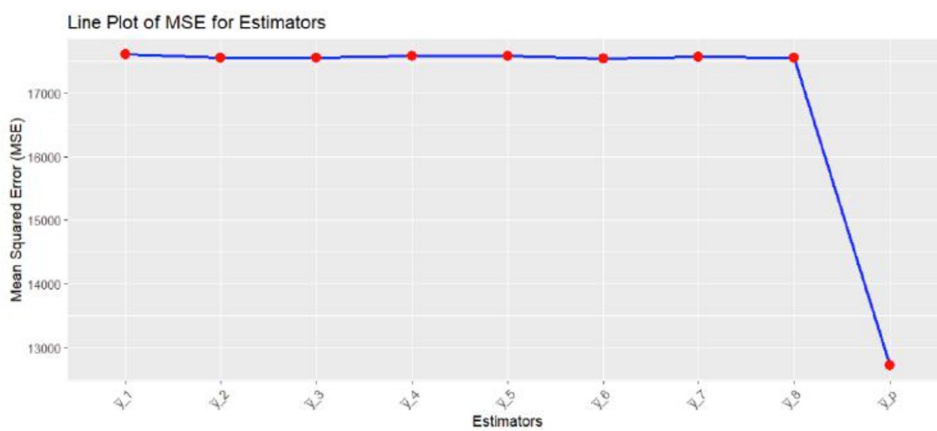
## Numerical Study

We have considered the population provided in [19] to evaluate the performance of the proposed modified ratio estimators with the other ratio estimators covered in this article. Mainly, the numerical study focuses on the mean estimation of loans for non-real estate based on the auxiliary information provided by loans of real estate. We plot the line charts, Graphs **1** and **2**, based on the MSE and PRE values provided in Table **2**.

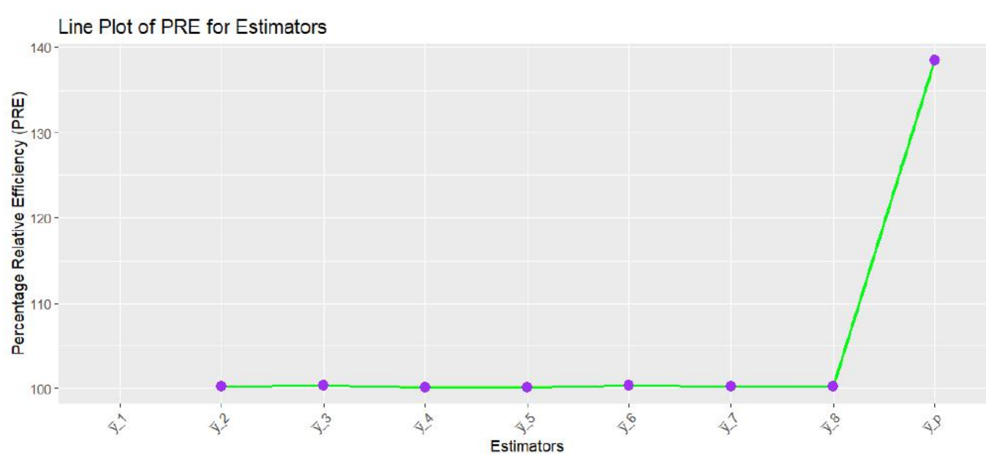
The formula for Percentage Relative Efficiency (PRE) is given by

**Table 2: The Values for the Constants $\theta$, Mean Square Error, Bias, and Percentage Relative Efficiency**

| Name of the Estimator | Values of $\theta$ | Bias | Mean Square Error | PRE |
|---|---|---|---|---|
| $\bar{y}_1$ | - | 28.00942 | 17606.43734 | - |
| $\bar{y}_2$ | 0.99860 | 27.84582 | 17562.97429 | 100.24747 |
| $\bar{y}_3$ | 0.99823 | 27.80264 | 17577.93 | 100.31289 |
| $\bar{y}_4$ | 0.99927 | 27.92407 | 17551.48 | 100.3105 |
| $\bar{y}_5$ | 0.99909 | 27.90304 | 17583.61 | 100.128 |
| $\bar{y}_6$ | 0.99817 | 27.79564 | 17538.65 | 100.32350 |
| $\bar{y}_7$ | 0.99864 | 27.85049 | 17564.21340 | 100.24040 |
| $\bar{y}_8$ | 0.99852 | 27.83648 | 17560.34 | 100.26161 |
| $\bar{y}_p$ | *0.69768 | *0.77515 | *12708.95 | *138.535 |




**Graph 1:** Line plot to represent the MSE values of estimators.




**Graph 2:** Line plot to represent the PRE values of estimators.

PRE $(\bar{y}_1, \bar{y}_i) = \frac{\mathrm{Var}(\bar{y}_1)}{MSE(\bar{y}_i)} \times 100$; i = 2,3,4,5,6,7,8, p.

Calculated numerical results of the estimators in this article are given in Table **2** and are calculated by using the constants provided in Table **1**.

## A Real Data Study

These days, a variety of cancers affect people. One of the aspects involved in determining the average number of fatalities from various cancer kinds is the estimation of deaths based on several auxiliary variables, such as the number of cases, tumor size, and survival. We present a real-world case study on datasets related to brain tumors [22] and breast cancer [21] in this article.

For dataset I, we used estimated deaths as the study variable (Y) and breast cancer cases as a related variable (X). Upon examining this dataset, we find a strong correlation between the impacted instances and patient fatalities. On the other hand, dataset II presents a realistic scenario about the number of fatalities from brain cancer (Y) according to tumor size (X).

## A Simulation Approach

The study's primary focus is on how the degree of correlation between the study & the associated variable improves population mean estimation. To improve the robustness of the suggested estimator, we tested its effectiveness by creating synthetic data with a sample size of 1000 drawn from a population of size (N) 100,000 and simulating this data 10,000 times. Tables

**Table 3: Required Data Statistics Analyzed from [21] and [22]**

| parameters | Data set I [21] | Data set II [22] |
|---|---|---|
| N | 500 | 1000 |
| n | 100 | 150 |
| $\bar{X}$ | 104141.8 | 1.553544 |
| $\bar{Y}$ | 19379.86 | 71.45935944 |
| $\rho$ | 0.8 | 0.031599032 |
| $s_x$ | 61070.46323 | 0.821487 |
| $M_d$ | 101361.5 | 1.572667 |
| $\beta_1$ | 0.070293804 | -0.03328 |
| $\beta_2$ | -1.09114146 | -1.10258 |
| $s_k$ | 0.13658 | -0.06983 |
| $C_x$ | 0.58641623 | 0.528782 |
| $C_y$ | 0.586430417 | 0.243945 |

**Table 4: Calculated MSE and PRE Values Based on Table 3**

| Name of the Estimator | Data set-I | | Data set-II | |
|---|---|---|---|---|
| | MSE values | PRE values | MSE values | PRE values |
| $\bar{y}_1$ | 413307.8 | 100 | 9.577047 | 100 |
| $\bar{y}_2$ | 413305.5 | 100.0006 | 6.049487 | 158.3117 |
| $\bar{y}_3$ | 413304.7 | 100.0008 | 9.262387 | 103.3972 |
| $\bar{y}_4$ | 413315.2 | 99.99821 | 71.51413 | 13.39183 |
| $\bar{y}_5$ | 413310 | 99.99948 | 18.31133 | 52.3012 |
| $\bar{y}_6$ | 413312.2 | 99.99895 | 96.93026 | 9.880349 |
| $\bar{y}_7$ | 413369.5 | 99.98509 | 1.727655 | 554.3381 |
| $\bar{y}_8$ | 413307.4 | 100.0001 | 10.26478 | 93.30008 |
| $\bar{y}_p$ | 371986 | 111.1084 | 1.720268 | 556.7184 |

**Table 5: Calculated MSE Values Based on Correlation Coefficients**

| Name of the Estimator | Values of MSE based on the correlation coefficient | | | | |
|---|---|---|---|---|---|
| | $\rho$=0.5 | $\rho$=0.6 | $\rho$=0.7 | $\rho$=0.8 | $\rho$=0.9 |
| $\bar{y}_1$ | 62.55599 | 53.26847 | 43.67183 | 32.85686 | 19.75593 |
| $\bar{y}_2$ | 62.41666 | 53.14143 | 43.55778 | 32.75951 | 19.68247 |
| $\bar{y}_3$ | 62.4355 | 53.13663 | 43.53389 | 32.72235 | 19.64175 |
| $\bar{y}_4$ | 61.81134 | 52.58909 | 43.06317 | 32.33785 | 19.36432 |
| $\bar{y}_5$ | 62.47854 | 53.19786 | 43.60842 | 32.80271 | 19.71508 |
| $\bar{y}_6$ | 62.55545 | 53.26811 | 43.67135 | 32.85656 | 19.75582 |
| $\bar{y}_7$ | 134539 | 1853.048 | 1318.678 | 2211.328 | 7825.543 |
| $\bar{y}_8$ | 62.55534 | 53.26811 | 43.67124 | 32.85655 | 19.75591 |
| $\bar{y}_p$ | 18.70694 | 16.63058 | 14.20892 | 11.28086 | 7.347708 |

**Table 6: Calculated PRE Values Based on Correlation Coefficients**

| Name of the Estimator | Values of MSE based on the correlation coefficient | | | | |
|---|---|---|---|---|---|
| | $\rho$=0.5 | $\rho$=0.6 | $\rho$=0.7 | $\rho$=0.8 | $\rho$=0.9 |
| $\bar{y}_1$ | 100 | 100 | 100 | 100 | 100 |
| $\bar{y}_2$ | 100.2232 | 100.2391 | 100.2618 | 100.2972 | 100.3732 |
| $\bar{y}_3$ | 100.193 | 100.2481 | 100.3169 | 100.4111 | 100.5813 |
| $\bar{y}_4$ | 101.2047 | 101.2919 | 101.4134 | 101.605 | 102.0223 |
| $\bar{y}_5$ | 100.124 | 100.1327 | 100.1454 | 100.1651 | 100.2072 |
| $\bar{y}_6$ | 100.0009 | 100.0007 | 100.0011 | 100.0009 | 100.0006 |
| $\bar{y}_7$ | 0.046497 | 2.874641 | 3.311789 | 1.485843 | 0.252454 |
| $\bar{y}_8$ | 100.001 | 100.0007 | 100.0014 | 100.0009 | 100.0001 |
| $\bar{y}_p$ | 334.3999 | 320.3043 | 307.355 | 291.262 | 268.872 |

**5** and **6** present the MSE results based on various correlation coefficients.

## CONCLUSIONS

We suggested a ratio-type estimator in Simple Random Sampling based on the median of an auxiliary variable and Karl Pearson's coefficient of correlation. Although there is bias in the suggested estimate, it is significantly less than that of other estimators. Descriptive statistics measures, such as the average, median, and standard deviation, serve as the foundation for the proposed estimate. Using the Mean Squared Error (MSE) equations and the biases of the proposed estimator and other estimators in this work, we calculate the MSE and PRE values. The MSE and PRE of the proposed estimate are compared with those of competing estimators in this study. By generating lower MSE and better PRE values, the proposed estimator satisfies the efficiency criteria listed in the section on efficiency comparisons.

We tested the proposed estimator, along with many conventional estimators, on a real estate loan dataset to see how it might work in practice. To establish its medicinal relevance, we conducted additional empirical analyses on brain tumor and breast cancer datasets. These applications demonstrate the estimator's ability to handle skewed distributions common in clinical and epidemiological data. To back up the theoretical findings, we ran a simulation study that demonstrated the suggested estimator's higher accuracy and efficiency under a variety of distributional scenarios. The results show that our estimator outperforms the alternatives tested, particularly in circumstances involving asymmetry or outliers. Given its ability to handle skewed data, we believe our estimator will be especially beneficial in medical research situations such as rare illness investigations, survival analysis, and biomarker-driven estimation, where traditional mean-based approaches may fall short. Moving forward, we propose additional study and potential collaborations with medical researchers to use our estimator in real-world healthcare applications such as clinical trials and public health surveys with small or stratified samples.

## CONFLICTS OF INTEREST

There is no conflict of interest.

## REFERENCES

[1] Sisodia BVS, Dwivedi VK. A modified ratio estimator using co-efficient of variation of auxiliary variable. Journal of the Indian Society of Agricultural Statistics 1981; 33: 13-18.

[2] Prasad B. Some improved ratio type estimators of population mean and ratio in finite population sample surveys. Communications in Statistics: Theory and Methods 1989; 18: 379-392. https://doi.org/10.1080/03610928908829905

[3] Upadhyaya LN, Singh HP. Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal 1999; 41: 627-636. https://doi.org/10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.0.CO;2-W

[4] Kadilar C, Cingi H. Ratio Estimators in Stratified Random Sampling. Biometrical Journal 2003; 45: 218-225. https://doi.org/10.1002/bimj.200390007

[5] Singh HP, Tailor R. Use of known correlation coefficient in estimating the finite population mean. Statistics in Transition 2003; 6: 555-560.

[6] Kadilar C, Cingi H. Ratio estimators in simple random sampling. Applied Mathematics and Computation 2004; 151: 893-902. https://doi.org/10.1016/S0096-3003(03)00803-8

[7] Singh HP, Tailor R, Kakran MS. An improved estimator of population Mean using power transformation. Journal of the Indian Society of Agricultural Statistics 2004; 58: 223-230.

[8] Singh HP, Tailor R. Estimation of finite population mean with known coefficient of variation of an auxiliary. Statistica 2005; 65: 301-313

[9] Kadilar C, Cingi H. An improvement in estimating the population mean by using the correlation coefficient. Hacettepe Journal of Mathematics and Statistics 2006; 35: 103-109.

[10] Koyuncu N, Kadilar C. Efficient Estimators for the Population Mean. Hacettepe Journal of Mathematics and Statistics 2009; 38: 217-225. https://doi.org/10.1080/03610920802562723

[11] Yan Z, Tian B. Ratio method to the mean estimation using co-efficient of skewness of auxiliary variable. Information computing and applications 2010; part II: 103-110. https://doi.org/10.1007/978-3-642-16339-5_14

[12] Bhushan S. Some efficient sampling strategies based on ratio type estimator. Electronic Journal of Applied Statistical Analysis 2012; 5: 74-88.

[13] Enang EI, Akpan VM, Ekpenyong EJ. Alternative ratio estimator of population Mean in simple random sampling. Journal of Mathematics Research 2014; 6: 54. https://doi.org/10.5539/jmr.v6n3p54

[14] Kosgey SC, Odongo L. Ratio estimator of population mean in simple random sampling. American Journal of Theoretical and Applied Statistics 2022; 13: 167-174.

[15] Ahmad S, Aamir M, Hussain S, Shabbir J, Zahid E, Subkrajang K, Jirawattanapanit A. A New Generalized Class of Exponential Factor-Type Estimators for Population Distribution Function Using Two Auxiliary Variables. Mathematical Problems in Engineering 2022; 2022: 2545517. https://doi.org/10.1155/2022/2545517

[16] Kumar M, Ashish Tiwari A. A Composite Class of Ratio Estimators for the Mean of a Finite Population in Simple Random Sampling. Advances and Applications in Mathematical Sciences 2023; 22: 2109-2123.

[17] Ounrittichai N, Utha P, Choopradit B, Chaipitak S. Performance Comparison of Three Ratio Estimators of the Population Ratio in Simple Random Sampling Without Replacement. International Journal of Analysis and Applications 2024; 22: 121-121. https://doi.org/10.28924/2291-8639-22-2024-121

[18] Cochran WG. Sampling Techniques, John Wiley and Son, New York 1977.

[19] Singh S. Advanced Sampling Theory with Applications: How Michael "Selected" Amy, Springer Science and Business Media 2003; Vol. 2. https://doi.org/10.1007/978-94-007-0789-4

[20] Lakshmi NV, Danish F, Alrasheedi M. Enhanced estimation of finite population mean via power and log-transformed ratio estimators using an auxiliary variable in solar radiation data. Journal of Radiation Research and Applied Sciences 2025; 18(2): 101379. https://doi.org/10.1016/j.jrras.2025.101379

[21] Gupta N. Global insights into breast cancer. Kaggle 2025. https://www.kaggle.com/code/nileshely/global-insights-into-breast-cancer

[22] Miadul ND. Brain Tumor Dataset [Data set]. Kaggle. Retrieved April 23, 2025, from https://www.kaggle.com/datasets/miadul/brain-tumor-dataset

[23] Lu J. Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural, biomedical, and power engineering. Mathematical Problems in Engineering 2017; 2017(1): 8704734. https://doi.org/10.1155/2017/8704734

[24] Kim JM, Tailor R, Sharma B. A generalized ratio-cum-product estimator of finite population mean in stratified random sampling. Communications for Statistical Applications and Methods 2011; 18(1): 111-118. https://doi.org/10.5351/CKSS.2007.14.1.111

Received on 17-05-2025 Accepted on 13-06-2025 Published on 17-07-2025

https://doi.org/10.6000/1929-6029.2025.14.36