IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Construction of an Almost Unbiased Estimator for Population Variance Using Exponential-Sine Type Estimator

Rajesh Singh[1], Sunil Kumar Yadav[1*]

[1]Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005, Uttar Pradesh, India.

[1]E-mail : rsinghstat@gmail.com
[*1]E-mail : ysunilkumar40@gmail.com

ARTICLE INFO

Article history:
Received 31 July 2025
Revised 01 August 2025,
Accepted 17 October 2025
Available online 04 November 2025

Keywords:
Population variance
Auxiliary information
Exponential - Sine estimator
Mean square error
Regression estimator

ABSTRACT

In this paper, we propose a generalised almost unbiased estimator for estimating the unknown finite population variance S_y^2 of the study variable Y, using auxiliary information within the exponential-cum-sine estimator framework. In many applications, bias in an estimator can be a significant drawback. Following the procedure of Singh, R. and Singh, S., we develop an estimator based on a single auxiliary variable, which is almost unbiased up to the order $O(n^{-1})$. Expressions for the bias and mean squared error (MSE) of the proposed estimator are derived up to the first order of approximation. To support the theoretical results, an empirical study using two real-life data sets has been conducted. Additionally, a simulation study confirms that the proposed estimator exhibits lower bias (nearly zero) and achieves a minimum MSE comparable to that of the regression estimator.

1. Introduction

In survey sampling, auxiliary information helps to make estimators more accurate and efficient. When no auxiliary information is available, the easiest way to estimate the population mean is by using the sample mean from simple random sampling without replacement. But if there is an auxiliary variable X that is closely related to the study variable Y, using ratio or regression estimators can give more accurate results. Estimating the finite population variance is of considerable importance across various domains as agriculture, industry, and the medical and biological sciences, particularly when dealing with skewed populations. Variability is a natural phenomenon; no two individuals or objects are exactly alike. For instance, in medical practice, understanding the variation in human blood pressure, body temperature, and pulse rate is crucial for accurate diagnosis and effective treatment. This paper focuses on estimating the population variance and demonstrates how incorporating auxiliary information can enhance the efficiency of variance estimation.

Isaki [1] was among the first to propose ratio and regression-based estimators for population variance. Later, Prasad and Singh [2] refined Isaki's estimator by reducing its bias and improving its precision. Arcos et al. [3] further extended this work by introducing an improved ratio-type estimator, which demonstrated lower bias and greater accuracy compared to Isaki's original approach and other existing estimators. The problem of constructing efficient estimators for a population variance has been discussed by many authors, including Upadhyaya and Singh [4], Singh et.al.[5], Das and Tripathi [6], Garcia and Cebrián [7], Gupta and Shabbir [8], Kadilar and Cingi [9,10], Singh et. al [11], Mishra and Singh [12,13], Singh and Yadav [14], Audu et.al.[15], Singh and Rai [16], Sharma and Singh [17,18], Singh et.al.[19-22] and many more scholars, such as Subramani and Kumarapandiyan [23-25], and more recently Adichwal et al. [26,27], have highlighted the vital role of variance estimation in survey sampling.

This study proposes improved exponential sine-type estimators for estimating the population variance using

Corresponding author E-mail address: ysunilkumar40@gmail.com
https://doi.org/10.62933/mz0yvj11

known auxiliary information under simple random sampling. To the best of our knowledge, this is the first application of exponential sine-type estimators in the context of variance estimation. Previous studies by Bhattacharyya et al. [28] and Yunusa et al. [29] applied sine-type methods for estimating the population median and mean; however, these methods have not been extended to variance estimation. To achieve this, we developed new exponential sine-type estimators by modifying existing variance estimators, including Isaki's [1] estimators and the exponential ratio and product-type estimators introduced by Bahl and Tuteja [30].

The estimator is constructed using two well-known functions, the exponential and sine functions, due to their usefulness in real-life applications. The exponential function helps model things that grow or decrease quickly, like how a disease spreads or how money grows with interest. The sine function is used for things that are periodic in nature, like sound waves, ocean tides, light waves, or changes in temperature during the year.

After proposing exponential sine-type estimators, we used a method by Singh and Singh [31,32] to make an almost unbiased estimator. This approach significantly reduces bias, resulting in estimates that are much closer to the true population variance. The primary goal of this study is to enhance the accuracy and minimise the error in estimating the variance by effectively incorporating auxiliary information.

1.1 Notation and terminology

Let Y and X be the study and auxiliary variables, respectively, of the population of interest, each measured for all population units U_i $(U_1, U_2, \dots\dots\dots U_N)$. A sample of size n is selected from this population using simple random sampling without replacement. y_i and x_i are the study and auxiliary variables corresponding i^{th} sample unit selected. The population means of these variables are $\bar{Y}$ and $\bar{X}$, respectively. The variances of the study and auxiliary variables are represented by S_y^2 and S_x^2 for the population, and s_y^2 and s_x^2 for the sample.

$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ are the sample means.

$S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$, $S_x^2 = \sum_{i=1}^{N}(x - \bar{X})^2$ are the population variances.

$s_y^2 = \frac{1}{N}\sum_{i=1}^{n}(y_i - \bar{Y})^2$, $s_x^2 = \sum_{i=1}^{n}(x - \bar{X})^2$, are the sample variances.

These notations are used to derive the bias and MSE of estimators.

$\theta = \left(\frac{1}{n} - \frac{1}{N}\right)$, $g = \frac{n}{N-n}$

$e_0 = \frac{s_y^2 - S_y^2}{S_y^2}$, $e_1 = \frac{s_x^2 - S_x^2}{S_x^2}$, $e_2 = \frac{s_z^2 - S_z^2}{S_z^2}$,

$E(e_0) = E(e_1) = E(e_2) = 0$

$E(e_0^2) = \frac{1}{n}(\partial_{400} - 1)$, $E(e_1^2) = \frac{1}{n}(\partial_{040} - 1)$, $E(e_2^2) = \frac{1}{n}(\partial_{004} - 1)$,

$E(e_0 e_1) = \frac{1}{n}(\partial_{220} - 1)$,

$E(e_0 e_2) = \frac{1}{n}(\partial_{202} - 1)$,

$E(e_1 e_2) = \frac{1}{n}(\partial_{022} - 1)$,

$\partial_{pqr} = \frac{\mu_{pqr}}{\mu_{200}^{\frac{p}{2}} \mu_{020}^{\frac{q}{2}} \mu_{002}^{\frac{r}{2}}}$, and

$\mu_{pqr} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^p (x_i - \bar{X})^q (z_i - \bar{Z})^r$,

2.Existing Estimator

This section reviews some commonly used estimators for the estimation of unknown population variance. The usual estimator for the population variance is defined as

$$t_0 = s_y^2 \tag{2.1}$$

The expressions for the bias and MSE of the usual estimator of population variance are derived as follows:

$$Bias(t_0) = 0 \tag{2.2}$$

$$MSE(t_0) = S_y^4 \, \gamma \, (\partial_{400} - 1) \tag{2.3}$$

Here $\gamma = \left(\frac{1}{n} - \frac{1}{N}\right)$

A generalised ratio-type estimator proposed by Isaki [1] is given as

$$t_1^k = s_y^2 \left(\frac{S_x^2}{s_x^2}\right)^k \tag{2.4}$$

The expressions for the bias and MSE of the estimator t_1^k are given as follows:

$$Bias(t_1^k) = S_y^4 \gamma \begin{bmatrix} \frac{k(k+1)}{2}(\partial_{040} - 1) \\ -k(\partial_{220} - 1) \end{bmatrix} \tag{2.5}$$

$$MSE(t_1^k) = S_y^4 \, \gamma [(\partial_{400} - 1) + k^2(\partial_{040} - 1) - 2k(\partial_{220} - 1)] \tag{2.6}$$

By minimising expression (2.6) with respect to k, we obtain the optimal value of k as:

$$k_{opt} = \frac{\partial_{220} - 1}{\partial_{040} - 1} \tag{2.7}$$

Substituting this value into equation (2.6), we obtain the minimum MSE of the estimator t_1^k as follows:

$$Min.MSE(t_1^k) = S_y^4 \gamma \left[(\partial_{400} - 1) - \left(\frac{(\partial_{220} - 1)^2}{\partial_{040} - 1}\right)\right] \tag{2.8}$$

For $k = 1$, the estimator reduces to the simple ratio estimator for population variance suggested by Isaki[1], given by:

$$t_{1r} = s_y^2 \left(\frac{S_x^2}{s_x^2}\right) \tag{2.9}$$

The bias and MSE of the estimator t_{1r} are respectively given as

$$Bias(t_{1r}) = S_y^2 \gamma [\partial_{040} - \partial_{220}] \tag{2.10}$$

$$MSE(t_{1r}) = S_y^4 \gamma [\partial_{400} + \partial_{040} - 2\partial_{220}] \tag{2.11}$$

For k = -1 the estimators reduce to the simple product type estimator for variance as:

$$t_{1p} = s_y^2 \left(\frac{s_x^2}{S_x^2}\right) \tag{2.12}$$

The bias and MSE for the estimator t_{1p} are respectively given by

$$Bias(t_{1p}) = S_y^2 \gamma (\partial_{220} - 1) \tag{2.13}$$

$$MSE(t_{1p}) = S_y^4 \gamma \begin{bmatrix} \partial_{400} + \partial_{040} \\ +\partial_{220} - 4 \end{bmatrix} \tag{2.14}$$

Inspired by the method as suggested by Bahl and Tuteja [30], an innovative generalised form of exponential-cum-sine type ratio estimator is constructed to estimate the population variance as follows

$$t_2^k = s_y^2 \exp\left[sin\left\{\frac{k(S_x^2 - s_x^2)}{S_x^2 + s_x^2}\right\}\right] \tag{2.15}$$

Up to the first order of approximation, the bias and MSE of the estimator t_2^k are respectively expressed as follows:

$$Bias(t_2^k) = S_y^2\gamma\begin{bmatrix}\left(\frac{k}{4}+\frac{k^2}{8}\right)(\partial_{040}-1)\\ -\frac{k}{2}(\partial_{220}-1)\end{bmatrix} \quad (2.16)$$

$$MSE(t_2^k) = S_y^4\,\gamma\begin{bmatrix}(\partial_{400}-1)+\\ \frac{k^2}{4}(\partial_{040}-1)-k(\partial_{220}-1)\end{bmatrix} \quad (2.17)$$

By minimising the estimator with respect to k, we obtain the optimal value of k as k_{opt}:

$$k_{opt} = \frac{2(\partial_{220}-1)}{\partial_{040}-1} \quad (2.18)$$

Substituting this value in equation (2.17), we get the minimum MSE of the estimator t_2^k as

$$Min.MSE(t_2^k) = S_y^4\gamma\left[(\partial_{400}-1)-\left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right] \quad (2.19)$$

For the particular case of $k=1$, the estimator reduces to exponential-cum-sine type ratio estimator is defined below:

$$t_{2r} = s_y^2\exp\left[sin\left(\frac{S_x^2-s_x^2}{S_x^2-s_x^2}\right)\right], \quad (2.20)$$

The bias and MSE of the estimator t_{2r}, up to the first-order approximation, are given as follows:

$$Bias(t_{2r}) = S_y^2\gamma\left[\frac{3}{8}(\partial_{040}-1)-\frac{1}{2}(\partial_{220}-1)\right], \quad (2.21)$$

$$MSE(t_{2r}) = S_y^4\gamma\begin{bmatrix}(\partial_{400}-1)+\\ \frac{1}{4}(\partial_{040}-1)-(\partial_{220}-1)\end{bmatrix}, \quad (2.22)$$

Similarly, for the case of $k=-1$, the estimators reduce exponential-cum-sine type product estimator and are defined as below:

$$t_{2p} = s_y^2\exp\left[sin\left(\frac{s_x^2-S_x^2}{S_x^2+s_x^2}\right)\right], \quad (2.23)$$

The bias and MSE of this product-type exponential-cum-sine estimator are expressed below, up to the first-order approximation.

$$Bias(t_{2p}) = S_y^2\gamma\left[\frac{1}{2}(\partial_{220}-1)-\frac{1}{8}(\partial_{040}-1)\right], \quad (2.24)$$

$$MSE(t_{2p}) = S_y^4\gamma\begin{bmatrix}(\partial_{400}-1)+\\ \frac{1}{4}(\partial_{040}-1)+(\partial_{220}-1)\end{bmatrix}, \quad (2.25)$$

3. Proposed Estimator using exponential sine estimator

Let $t_0 = s_y^2$, $t_1 = s_y^2\left(\frac{S_x^2}{s_x^2}\right)^k$ and $t_2 = s_y^2 exp\left[sin\left(\frac{k(S_x^2-s_x^2)}{S_x^2+s_x^2}\right)\right]$ are the three estimators. These estimators are used to construct a linear form of the estimator, with constants α_1, α_2 and α_3 as weights. Here t_0, t_1 and t_2 $\in$ L, where L denotes the set of all possible estimators for estimating the unknown population variance S_y^2.

Based on the work of Singh, R. and Singh, S. [31, 32], we have developed an almost unbiased estimator t_h by linearly combining three initial estimators t_0, t_1 and t_2 for the estimation of the population variance. By imposing a linear constraint $t_h = \sum_{i=0}^{2}\alpha_i t_i \in$L, we determine the weights α_1, α_2 and α_3 explicitly, based on the correlation and relative variance of the auxiliary variables. This approach reduces bias up to the first order of approximation and ensures higher efficiency, particularly when the study and auxiliary variables are correlated.

By definition,

$$t_h = \sum_{i=0}^{2}\alpha_i\, t_i \in L, \quad (3.1)$$

The extended form of the proposed estimator is -

$$t_h = \alpha_0 s_y^2 + \alpha_1\, s_y^2\left(\frac{S_x^2}{s_x^2}\right)^k + \alpha_2\, s_y^2\exp\left[sin\left(\frac{k(S_x^2-s_x^2)}{s_x^2+S_x^2}\right)\right], \quad (3.2)$$

For $\sum_{i=0}^{2}\alpha_i = 1$, $\alpha_i \in R$ (3.3)

where α_i $(i=0,1,2)$ denotes the statistical constants and R denotes the set of real numbers.

Table 1. For k =1, the corresponding member of the class of estimators

α_0	α_1	α_2	**Estimators**
1	0	0	s_y^2
0	1	0	$s_y^2\left(\frac{S_x^2}{s_x^2}\right)$
0	0	1	$s_y^2\exp\left[sin\left(\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right)\right]$

Table 2. For $k = -1$, the corresponding member of the class of estimators

α_0	α_1	α_2	Estimators
1	0	0	s_y^2
0	1	0	$s_y^2\left(\frac{s_x^2}{S_x^2}\right)$
0	0	1	$s_y^2 \exp\left[sin\left(\frac{s_x^2 - S_x^2}{S_x^2 + s_x^2}\right)\right]$

When Equation (3.2) is expanded using a first-order approximation in terms of the error components (e), the resulting form is given by:

$$t_h = S_y^2\left[\alpha_0(1+e_0) + \alpha_1(1+e_0)\left(\frac{1}{1+e_1}\right)^k + \alpha_2(1+e_0)exp\left(\sin\left(\frac{-k\,e_1}{2+e_1}\right)\right)\right], \quad (3.4)$$

By expanding this equation, we obtain the following expression:

$$t_h = S_y^2\left[\alpha_0(1+e_0) + \alpha_1(1+e_0)(1-ke_1+De_1^2) + \alpha_2(1+e_0)\left(1-\frac{k}{2}e_1+Ce_1^2\right)\right], \quad (3.5)$$

where,

$D = \frac{k(k+1)}{2}$, $\quad C = \left(\frac{k}{4}+\frac{k^2}{8}\right)$.

After simplification, the resulting expression is:

$$t_h = S_y^2\left[(\alpha_0+\alpha_1+\alpha_2) + (k_0+k_1+k_2)e_0 - k\alpha_1e_1 - \frac{1}{2}k\alpha_2e_1 - k\alpha_1e_0e_1 - \frac{1}{2}k\alpha_2e_0e_1 + D\alpha_1e_1^2 + C\alpha_2e_1^2\right], \quad (3.6)$$

The above expression is equivalent to:

$$t_h = S_y^2\left[1+e_0-\left(\alpha_1+\frac{1}{2}\alpha_2\right)ke_1 - \left(\alpha_1+\frac{1}{2}\alpha_2\right)ke_0e_1 + (D\alpha_1+C\alpha_2)e_1^2\right], \quad (3.7)$$

where,

$\alpha_0+\alpha_1+\alpha_2 = 1$ and $H = (\alpha_1+\frac{1}{2}\alpha_2)$

Subtracting $\bar{Y}$ from both sides of Equation (3.7) and then taking the expectation, we derive the bias of the estimator t_h as follows:

$$Bias(t_h) = S_y^2\gamma[(D\alpha_1+C\alpha_2)(\partial_{040}-1) - Hk(\partial_{220}-1)], \quad (3.8)$$

The equation (3.8) can be written as:

$$(t_h - \bar{Y}) \cong S_y^2[e_0 - Hke_1], \quad (3.9)$$

Squaring both sides of Equation (3.9) and taking the expectation, the MSE of the estimator t_h is obtained as:

$$MSE(t_h) = S_y^2\gamma[(\partial_{400}-1) + H^2k^2(\partial_{040}-1) - 2Hk(\partial_{220}-1)], \quad (3.10)$$

By minimising this expression with respect to H, we obtain the optimal value of H as follows:

$$H_{opt} = \left(\alpha_1+\frac{1}{2}\alpha_2\right) = \frac{\partial_{220}-1}{k(\partial_{040}-1)} \quad (3.11)$$

Substituting this optimal value of H_{opt} into Equation (4.10), we obtain the minimum MSE of the estimator t_h as follows:

$$Min.MSE(t_h) = S_y^4\gamma\left[(\partial_{400}-1) - \left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right] \quad (3.12)$$

It is observed that the minimum MSE of the estimator t_h is equivalent to the MSE of the regression-type estimator when estimating the population variance.

Equations (3.3) and (3.11) together constitute a system with only two equations involving three unknown parameters α_i, which makes the system incomplete and prevents us from finding unique values for these coefficients. To address this issue, we impose the following linear restriction to obtain a unique solution:

$$\sum_{i=1}^{2}\alpha_i\,B(t_i) = 0, \quad (3.13)$$

Such that

$$\alpha_0B(t_0) + \alpha_1B(t_1) + \alpha_2B(t_2) = 0\,, \quad (3.14)$$

Here $B(t_i)$ presents the bias of the i^{th} $i = 0,1,2$ estimator.

The equation is written in matrix form as

$$\begin{bmatrix}1 & 1 & 1\\ 0 & 1 & \frac{1}{2}\\ 0 & B(t_1) & B(t_2)\end{bmatrix}\begin{bmatrix}\alpha_0\\ \alpha_1\\ \alpha_2\end{bmatrix} = \begin{bmatrix}1\\ H\\ 0\end{bmatrix} \quad (3.15)$$

where (t_0) , $B(t_1)$ and $B(t_2)$ are defined in equation (2.2),(2.10) and (2.21).

From the system of equations, we get the unique values of $\alpha_i{}'s(i = 0,1\ and\ 2)$.

$$\alpha_0 = \frac{\left[\left(B(t_2)-\frac{1}{2}B(t_1)\right)-HB(t_2)+HB(t_1)\right]}{B(t_2)-\frac{1}{2}B(t_1)}, \quad (3.16)$$

$$\alpha_1 = \frac{HB(t_2)}{\left(B(t_2)-\frac{1}{2}B(t_1)\right)}, \quad (3.17)$$

$$\alpha_2 = \left(-\frac{HB(t_1)}{\left(B(t_2)-\frac{1}{2}B(t_1)\right)}\right), \quad (3.18)$$

where,

$$\alpha_0 + \alpha_1 + \alpha_2 = 1 \quad (3.19)$$

The use of these $\alpha_i's$ ($i = 0,1\ and\ 2$) coefficients eliminates the bias of the proposed estimator up to the first-order approximation.

4. Theoretical Comparison of Min. MSE with other existing estimators

1. Estimator t_h will be more efficient than t_0 if,

$$Min.MSE(t_h) - MSE(t_0) < 0$$

$$\left[(\partial_{400}-1)-\left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right]-(\partial_{400}-1)<0 \qquad (4.1)$$

2. Estimator t_h will be more efficient than t_1 if,

$$Min.MSE(t_h) - MSE(t_1) < 0$$

$$\left[(\partial_{400}-1)-\left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right]-[\partial_{400}+\partial_{040}-2\partial_{220}]<0 \qquad (4.2)$$

3. Estimator t_h will be more efficient than t_2 if,

$$Min.MSE(t_h) - MSE(t_2) < 0$$

$$\left[(\partial_{400}-1)-\left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right]-[\partial_{400}+\partial_{040}+\partial_{220}-4]<0 \qquad (4.3)$$

4. Estimator t_h will be more efficient than t_3 if,

$$Min.MSE(t_h) - MSE(t_3) < 0$$

$$\left[(\partial_{400}-1)-\left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right]-\left[(\partial_{400}-1)+\frac{1}{4}(\partial_{040}-1)-(\partial_{220}-1)\right]<0 \qquad (4.4)$$

5. Estimator t_h will be more efficient than t_4 if,

$$Min.MSE(t_h) - MSE(t_4) < 0$$

$$\left[(\partial_{400}-1)-\left(\frac{(\partial_{220}-1)^2}{\partial_{040}-1}\right)\right]-\left[(\partial_{400}-1)+\frac{1}{4}(\partial_{040}-1)+(\partial_{220}-1)\right]<0 \qquad (4.5)$$

5. Empirical study

To demonstrate the efficiency of different estimators of the population variance , we use the following datasets.

Population 1. (Source: Singh [33]),
y: fish caught in the year 1995,
x: fish caught in the year 1993,
z: fish caught in the year 1994.
$N = 69,\ n = 25$, ∂_{400} =7.7685 , ∂_{040} = 9.986 , ∂_{004} = 9.9851, ∂_{220} = 8.3107, ∂_{202} = 8.1715, ∂_{022} = 9.6631, S_y^2 = 37199578,

Population 2.
(Source: Murthy [34], p.399).
y: area under wheat in 1964,
x: area under wheat in 1963,
z: cultivated area in 1961.
$N = 34,\ n = 15$, ∂_{400} =3.7879 , ∂_{040} = 2.9123, ∂_{004} = 2.8082, ∂_{220} = 3.1046, ∂_{202} = 2.979, ∂_{022} = 2.7379, S_y^2 = 22564.55704,

Table 3: Values of $\alpha_i\ (i = 0, 1, 2)$

Scalars	Population 1 Values	Population 2 Values
α_0	-0.4203	0.3421
α_1	0.2069	1.5432
α_2	1.2134	-0.8854
$H\left(\alpha+\frac{1}{2}\alpha_2\right)$	0.8136	1.1006

By using these values of $\boldsymbol{\alpha_i}$ $(\boldsymbol{i = 0, 1, 2})$ given in Table 3, one can reduce the bias up to the order $O(n^{-1})$ in the estimator t_h.

Table 4: The Bias, MSE and PRE of the existing and proposed estimators in case of population 1.

Estimators	Bias	MSE	PRE
t_0	0	0.2707	100
t_{1r}	0.0670	0.0453	597.3436
t_{1p}	0.2924	0.8826	30.6750
t_{2r}	-0.0114	0.0682	397.1425
t_{2p}	0.1013	0.6530	41.4592
t_{hmin}	0	0.0328	824.6561

Table 5: The Bias, MSE and PRE of the existing and proposed estimators in case of population 2.

Estimators	Bias	MSE	PRE
t_0	0	0.1859	100
t_{1r}	-0.0128	0.0327	567.8004
t_{1p}	0.3870	0.3870	48.0275
t_{2r}	-0.0223	0.0774	240.0517

t_{2p}	0.0542	0.3580	51.9106
t_{hmin}	0	0.0314	591.0796

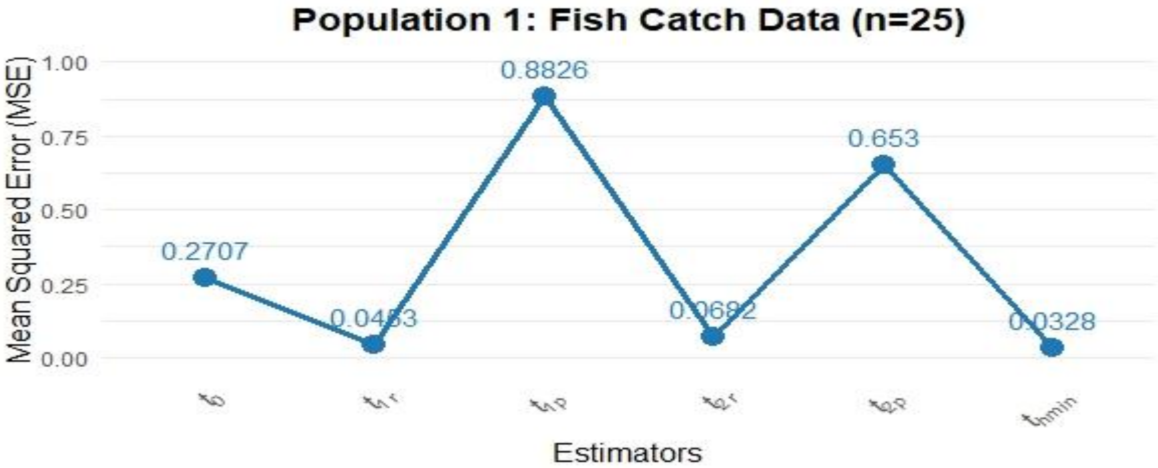

Figure 1. Line diagram of MSE across different estimators

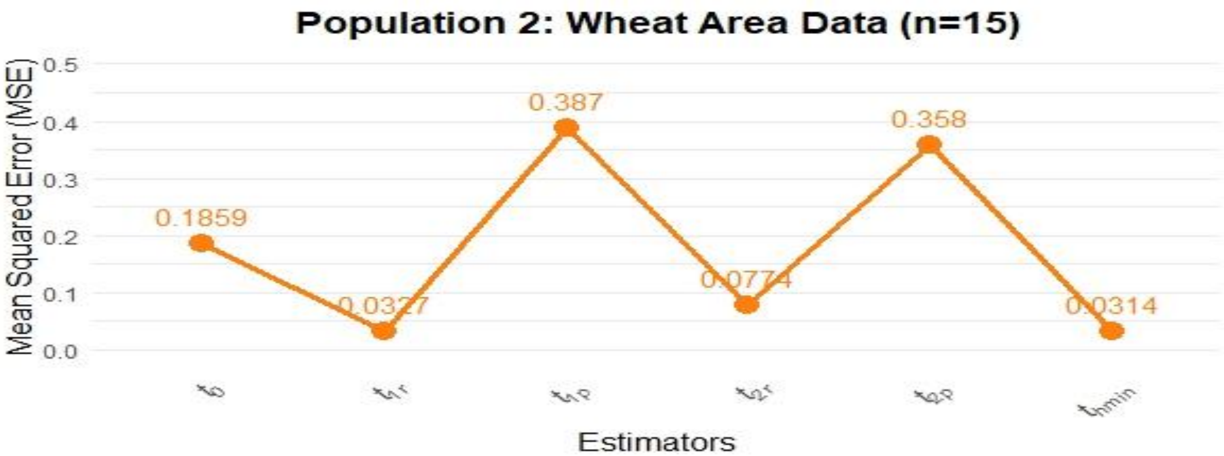

Figure 2. Line diagram of MSE across different estimators

6. Simulation Study

In this section, a simulation exercise has been conducted to evaluate the MSE and percentage relative efficiency (PRE) of the proposed estimators.

The following steps have been used for the simulation:

A bivariate normal population of size N=1000 is generated with the following parameters:

μ_x = 9, σ_x = 4;

μ_y = 25, σ_y = 9;

and correlation coefficients ρ = 0.75, 0.80, and 0.90.

A sample of sizes n = 100, 200 and 400 is selected from the simulated population.

For each sample, the sample mean, sample variance, and values of both the proposed and existing estimators of the population mean are calculated.

Steps 2 and 3 are repeated m= 10,000 times to ensure stability and accuracy of the results.

The mean squared error (MSE) of each estimator is computed based on the repeated simulations.

The percentage relative efficiency of each estimator t_h relative to a reference estimator t_0 is obtained using the following formula:

$$PRE(t_i, t_0) = \frac{MSE(t_0)}{MSE(t_i)} * 100$$

Table 6: Values of $\alpha_i\ (i = 0, 1, 2)$

N	Scalars	(N= 1000, ρ =0.75)	(N= 1000, ρ =0.80)	(N=1000, ρ=0.90)
100	α_0	-0.4595	-0.4776	-0.3732
	α_1	-0.2813	-0.1539	0.2701
	α_2	1.7408	1.6315	1.1032
	$H\left(\alpha + \frac{1}{2}\alpha_2\right)$	0.5891	0.6618	0.8217
200	α_0	-0.5135	-0.5174	-0.3822
	α_1	-0.3216	-0.1782	0.2733
	α_2	1.8351	1.6956	1.1089
	$H\left(\alpha + \frac{1}{2}\alpha_2\right)$	0.5960	0.6696	0.8278
400	α_0	-0.5426	-0.5386	-0.3867
	α_1	-0.3423	-0.1903	0.2753
	α_2	1.8849	1.7290	1.1114
	$H\left(\alpha + \frac{1}{2}\alpha_2\right)$	0.6001	0.6742	0.8310

By using these values of $\boldsymbol{\alpha_i}\ (\boldsymbol{i = 0, 1, 2})$ given in the table, one can reduce the bias up to the order $O(n^{-1})$ in the estimator t_h.

Table 7: The Bias, MSE and PRE of the existing and proposed estimators

N = 1000					
ρ	n	Estimators	Bias	MSE	PRE
0.75	100	t_0	0	0.0199	100
		t_{1r}	0.0076	0.0164	121.2251
		t_{1p}	0.0111	0.0397	50.1839
		t_{2r}	0.0015	0.0135	147.5004
		t_{2p}	0.0032	0.0357	55.8341
		t_{hmin}	-0.0035	0.0129	154.6311
	200	t_0	0	0.0101	100
		t_{1r}	0.0038	0.0083	122.4190
		t_{1p}	0.0057	0.0202	49.9989
		t_{2r}	0.0007	0.0068	148.2542
		t_{2p}	0.0016	0.0181	55.7626
		t_{hmin}	-0.0018	0.0066	152.6823
	400	t_0	0	0.0051	100
		t_{1r}	0.0019	0.0041	123.2528
		t_{1p}	0.0029	0.0102	49.8705
		t_{2r}	0.0004	0.0034	148.8137
		t_{2p}	0.0008	0.0091	55.6980
		t_{hmin}	-0.0010	0.0034	151.8811

Table 8: The Bias, MSE and PRE of the existing and proposed estimators

N = 1000					
ρ	**n**	**Estimators**	**Bias**	**MSE**	**PRE**
0.8	100	t_0	0	0.0199	100
		t_{1r}	0.0063	0.0136	146.7346
		t_{1p}	0.0126	0.0415	48.0466
		t_{2r}	0.0008	0.0120	165.7177
		t_{2p}	0.0040	0.0373	53.4418
		t_{hmin}	-0.0063	0.0111	179.2966
	200	t_0	0	0.0101	100
		t_{1r}	0.0031	0.0068	148.4886
		t_{1p}	0.0064	0.0211	47.8709
		t_{2r}	0.0004	0.0061	166.6810
		t_{2p}	0.0020	0.0189	53.3730
		t_{hmin}	-0.0033	0.0057	177.5934
	400	t_0	0	0.0051	100
		t_{1r}	0.0016	0.0034	149.6738
		t_{1p}	0.0033	0.0107	47.7459
		t_{2r}	0.0002	0.0030	167.3797
		t_{2p}	0.0010	0.0095	53.3109
		t_{hmin}	-0.0017	0.0029	176.9835

Table 9: The Bias, MSE and PRE of the existing and proposed estimators

N = 1000					
ρ	**n**	**Estimators**	**Bias**	**MSE**	**PRE**
0.9	100	t_0	0	0.0199	100
		t_{1r}	0.0034	0.0072	274.9647
		t_{1p}	0.0160	0.0453	43.9461
		t_{2r}	-0.0007	0.0087	227.9783
		t_{2p}	0.0056	0.0408	48.8336
		t_{hmin}	-0.0127	0.0065	306.5517
	200	t_0	0	0.0101	100
		t_{1r}	0.0017	0.0036	279.4535
		t_{1p}	0.0082	0.0231	43.7834
		t_{2r}	-0.0004	0.0044	229.3827
		t_{2p}	0.0028	0.0207	48.7809
		t_{hmin}	-0.0065	0.0033	306.6770
	400	t_0	0	0.0051	100
		t_{1r}	0.0008	0.0018	282.3322
		t_{1p}	0.0041	0.0117	43.6671
		t_{2r}	-0.0002	0.0022	230.3814
		t_{2p}	0.0014	0.0104	48.7328
		t_{hmin}	-0.0033	0.0017	307.3321

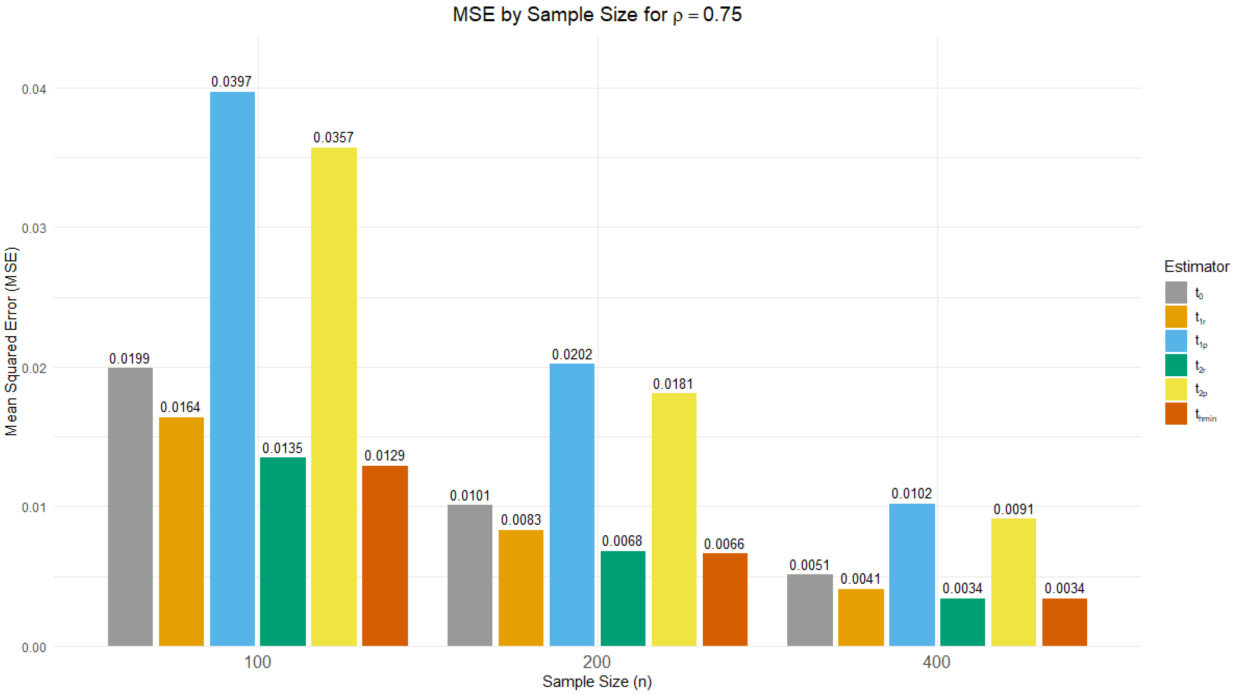

Figure 3. Bar graph of MSE by Sample Size at Correlation ρ = 0.75

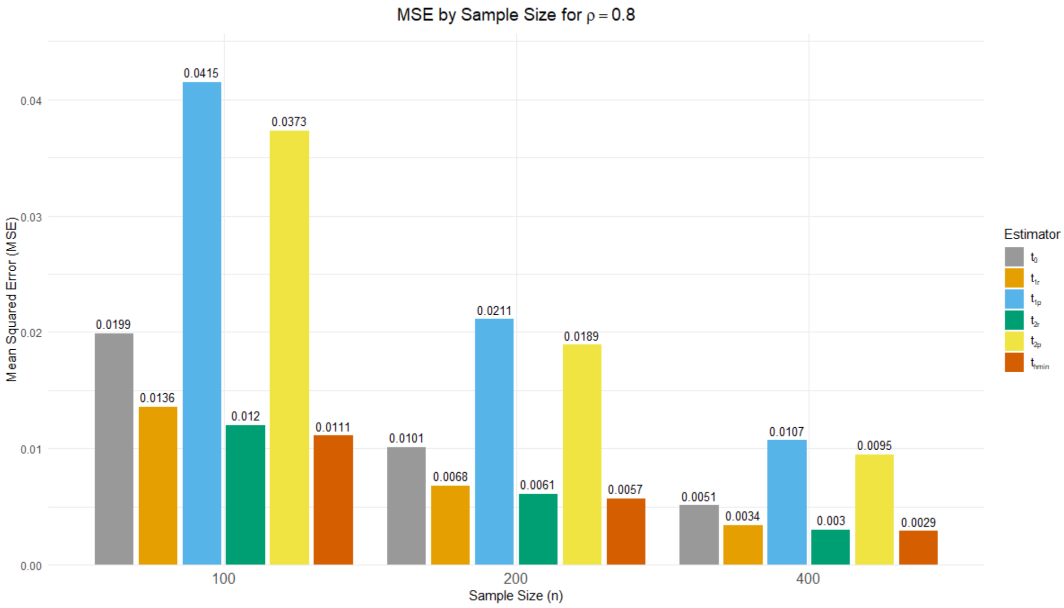

Figure 4. Bar graph of MSE by Sample Size at Correlation ρ = 0.80

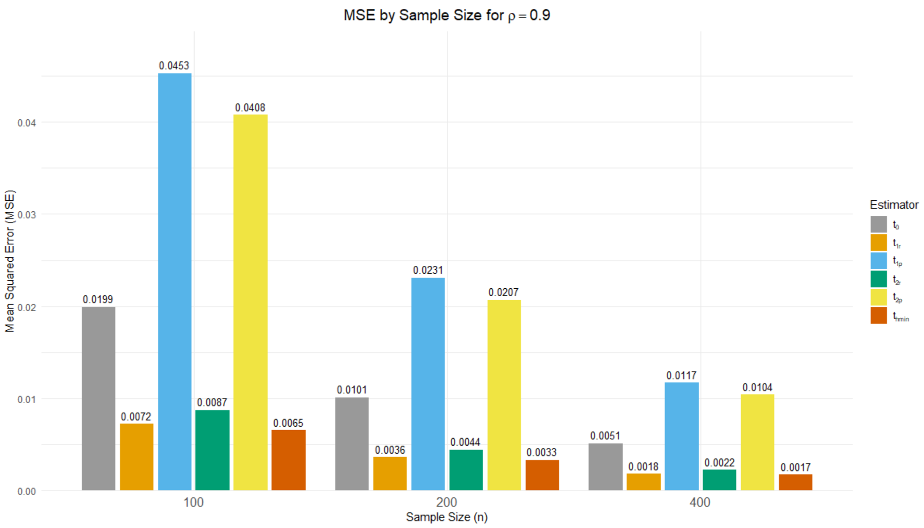

Figure 5. Bar graph of MSE by Sample Size at Correlation ρ = 0.90

7. Results and Discussion

This study evaluates the performance of the proposed estimator t_h and also compares it with existing estimators. For this study, we have considered two real data sets as population 1 and population 2 and also a simulation study. The estimators under comparison include the conventional estimator t_0, existing estimators t_{1r},t_{1p}, t_{2r}, t_{2p} and a newly proposed estimator denoted by t_h. The assessment is based on Bias, MSE, and PRE. Table 3 presents the values of the scalar constants α_0, α_1 and α_2 for both populations. These constants are incorporated into the construction of the proposed estimator t_h to reduce bias up to the first order of approximation.

In Tables 4 and 5, the empirical results for Population 1 (fish catch data) and Population 2 (wheat area data) are displayed. The estimator t_0 is taken as the baseline estimator with PRE 100. In both populations, the proposed estimator t_{hmin} exhibits the lowest MSE and the highest PRE (824.65 in the case of Population 1 and 591.08 in the case of Population 2), clearly establishing its superiority. The estimators t_{1r} and t_{2r} also perform significantly better than other product-type estimators t_{1p} *and* t_{2p}.

To further support these findings, a simulation study is conducted. A bivariate normal population of size N=1000 is generated with correlation coefficients (ρ = 0.75, 0.80, and 0.90). A Sample of sizes 100, 200, and 400 is repeatedly drawn to ensure robust and stable results. The corresponding values of α_i are calculated for each configuration and are reported in Table 6.

Tables 7, 8, and 9 summarise the performance of the estimators across different correlation values. In every case, the estimator t_{hmin} consistently achieves the smallest MSE and the highest PRE. This indicates superior efficiency in comparison to the existing estimators.

The line graphs show the MSE for different estimators in two cases: Population 1 (fish catch data) and Population 2 (wheat area data). Both populations present a similar pattern in how the estimators perform, but the actual MSE values are different. In Population 1, the MSE values are higher, and Population 2 shows that the MSEs are lower overall. The bar graphs show the MSE of estimators at different sample sizes (n = 100, 200, 400) and correlation levels (ρ = 0.75, 0.80, 0.90). One clear pattern is that MSE decreases as the sample size increases. This means that estimators become more accurate when more information is available. These visualisations confirm that statistical efficiency is strongly influenced by both the structure of the estimator and the quality of the data. A well-designed estimator, like t_{hmin}, consistently yields lower MSE, especially when supported by larger sample sizes and stronger correlations.

8.Conclusion

This paper introduces a new class of sine-type estimators for estimating the population variance under the simple random sampling without replacement (SRSWOR) scheme, utilising known auxiliary information. By incorporating exponential and sine transformations into traditional exponential ratio and product-type estimators, the proposed methods significantly enhance estimation accuracy. Additionally, an almost unbiased estimator is proposed. This estimator is constructed as a linear combination of selected exponential-cum-sine estimators, effectively reducing bias up to the first-order approximation.

Theoretical derivations of bias and MSE demonstrate that the proposed estimators outperform existing ones. These findings are further supported by empirical analyses using real datasets and a simulation study, which is conducted across various sample sizes and correlation levels. Among the proposed estimators, the almost unbiased estimator t_h consistently achieves the lowest MSE and the highest PRE, highlighting its strong practical utility. Overall, the study confirms that sine-exponential estimators, particularly the almost unbiased estimator, provide substantial efficiency improvements, especially when there is a strong correlation between the auxiliary and study variables, making them valuable tools for variance estimation in survey sampling.

References

[1] Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, *78*(381), 117–123.

[2] Prasad, B., & Singh, H. P. (1990). Some improved ratio type estimators of finite population variance in sample surveys. Communications in Statistics - Theory and Methods, 19(4), 1127–1139.

[3] Arcos, A., Barrios, E., & Sarabia, J. M. (2005). Incorporating the auxiliary information available in variance estimation. Applied Mathematics and Computation, 160(2), 387–399.

[4] Upadhyaya, L. N., & Singh, H. P. (2006). Almost unbiased ratio and product-type estimators of finite population variance in sample surveys. *Statistics in Transition*, *7*(5), 1087–1096.

[5] Singh, H. P., Chandra, P., & Singh, S. (2003). Variance estimation using multiauxiliary information for random non-response in survey sampling. *Statistica*, *63*(1), 23–40.

[6] Das, A. K., & Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhyā: Indian Journal of Statistics, Series B*, *40*, 139–148.

[7] Garcia, M. K., & Cebrián, A. A. (1997). Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator. *Metrika*, *45*, 171–178.

[8] Gupta, S., & Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, *37*, 57–67.

[9] Kadilar, C., & Cingi, H. (2006). Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, *35*(1), 111–115.

[10] Kadilar, C., & Cingi, H. (2006). Ratio estimators for population variance in simple and stratified sampling. *Applied Mathematics and Computation*, *173*, 1047–1058.

[11] Singh, J., Singh, V. K., Malik, S., & Singh, R. (2013). Use of auxiliary information in variance estimation. *Inspira Journal of Modern Management & Entrepreneurship*, *3*(1), 75–81.

[12] Mishra, P., & Singh, R. (2016). Variance estimation using arithmetic mean, geometric mean, and harmonic mean under simple random sampling. *Journal of Scientific Research*, *60*(1–2), 97–109.

[13] Mishra, P., & Singh, R. (2016). Variance estimation using quality characteristic. *Merit Research Journal of Interdisciplinary Studies*, *5*(2), 143–149.

[14] Singh, R., & Yadav, S. K. (2024). Efficient estimators of population coefficient of variation under simple random sampling using single auxiliary variable. *Journal of Statistics Applications & Probability Letters*, *11*(3), 109–121.

[15] Audu, A., Singh, R., Khare, S., & Duran, N. S. (2022). Almost unbiased variance estimators under the simultaneous influence of non-response and measurement errors. *Journal of Statistical Theory and Practice*.

[16] Singh, R., & Rai, S. (2021). Estimating population variance using median as an auxiliary information. *Journal of Scientific Research*, *65*(5), 222–225.

[17] Sharma, P., & Singh, R. (2013). A generalized class of estimators for finite population variance in presence of measurement errors. *Journal of Modern Applied Statistical Methods*, *12*(2), 231–241.

[18] Sharma, P., & Singh, R. (2014). Improved dual to variance ratio type estimator for population variance. *Chilean Journal of Statistics*, *5*(2), 45–54.

[19] Singh, R., Kumar, M., Singh, A. K., & Smarandache, F. (2011). A family of estimators of population variance using information on auxiliary attribute. *In SAM ME SER*, 63.

[20] Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2008). Almost unbiased ratio and product type estimator of finite population variance using the knowledge of kurtosis of an auxiliary variable in sample surveys. *Octogon Mathematical Journal*, *16*(1), 123–130.

[21] Singh, R., Sharma, P., & Adewara, A. (2013). A general class of estimators for estimating population variance in presence of measurement errors. *Nigerian Journal of Pure and Applied Sciences*, *26*, 2394–2401.

[22] Singh, R., Gupta, S., & Khare, S. (2019). An exponential type estimator for finite population variance. *International Journal of Applied Statistics & Systems*, *5*(2), 469–473.

[23] Subramani, J., & Kumarapandiyan, G. (2012). Estimation of population mean using coefficient of variation and median of an auxiliary variable. *International Journal of Probability and Statistics*, *1*(4), 111–118.

[24] Subramani, J., & Kumarapandiyan, G. (2012). Variance estimation using median of the auxiliary variable. *International Journal of Probability and Statistics*, *1*(3), 36–40.

[25] Subramani, J., & Kumarapandiyan, G. (2012). Variance estimation using quartiles and their functions of the auxiliary variable. *International Journal of Statistics and Applications*, *2*(5), 67–72.

[26] Adichwal, N. K., Sharma, P., Verma, H. K., & Singh, R. (2015). Generalised class of estimators for estimating population variances using auxiliary attributes. *International Journal of Applied and Computational Mathematics*, *1*(3), 1–10.

[27] Adichwal, N. K., Sharma, P., & Singh, R. (2015). Estimation of finite population variances using auxiliary attribute in sample surveys. *Journal of Advanced Computing*, *4*(2), 88–100.

[28] Bhattacharyya, D., Singh, G. N., Jawa, T. M., Sayed-Ahmed, N., & Pandey, A. K. (2021). An exponential-cum-sine-type hybrid imputation technique for missing data. *Computational Intelligence and Neuroscience*, *2021*(1), Article ID 4845569. https://doi.org/10.1155/2021/4845569

[29] Yunusa, M. A., Muili, J. O., Audu, A., & Singh, R. V. K. (2023). A sine-type median-based estimator for the estimation of population mean. *Oriental Journal of Physical Sciences*, *8*(1), 21–26. https://doi.org/10.13005/OJPS08.01.05

[30] Bahl, S., & Tuteja, R. K. (1991). Ratio and product-type exponential estimators. *Journal of Information & Optimisation Sciences*, *12*(1), 159–164.

[31] Singh, S., & Singh, R. (1993). Almost filtration of bias precipitates: A new approach. *Journal of the Indian Society of Agricultural Statistics*, *45*(2), 214–218.

[32] Singh, S., & Singh, R. (1991). Almost filtration of bias precipitates: A new technique. *Aligarh Journal of Statistics*, *11*, 5–8.

[33] Singh, S. (2003). *Advanced Sampling Theory with Applications: How "Michael" Selected "Amy"* (Vol. 2). New York, NY: Springer Science & Business Media.

[34] Murthy, M. N. (1967). *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.