International Journal of Advance Interdisciplinary Research
Vol. 2, Special Issue 1, (Jan-March) 2026, pp. 33-45, e-ISSN: 3107-913X
DOI: https://doi.org/10.66095/ijair.2026.v2.s1.4

Efficient estimator of population mean in SR

Karan Singh[1], Namita Srivastava[2] and Gunjan Singh[3*]
[1]Student, Amity School of Applied Sciences, Amity University Uttar Pradesh, Lucknow
[2]Professor, St Jhon's College, Agra
[3*]Assistant Professor, Amity School of Applied Sciences, Amity University Uttar Pradesh, Lucknow

Abstract: By guaranteeing a non-zero likelihood of choosing a population member, Probability Sampling techniques reduces bias and improves generalizability. There are several techniques, such as cluster sampling, multistage sampling, sampling with stratification, systematic sampling, and simple random sampling. This study examines probability sampling, its importance, and comparative analysis using Mean Square Error, bias, and efficiency. Cluster and sampling with multiple stages are helpful for populations that are spread out geographically, results show that sampling with stratification is the most dependable.

Keyword: Bias, Mean Square Error, Cluster Sampling, Systematic Sampling, Stratified Sampling, Probability Sampling, and Simple Random Sampling

Introduction

A popular research technique that guarantees an impartial and equitable proportion of a population is probability sampling. Probability sampling reduces selection bias and enables researchers to draw reliable statistical inferences about the population by using randomization [1]. Achieving representative data that promotes generalization while preserving effectiveness and economy is the main objective [2]. SRS, stratified sampling, cluster sampling, systematic sampling, and multiple stages of sampling are some of the probability sampling techniques; each has advantages and disadvantages [3].

Objectives

1) To investigate the basic ideas behind probability sampling and its importance in statistical analysis.
2) To carry out a simulation study to evaluate how well various sampling techniques perform under various population configurations.
3) To decide on the best probability sampling method based on feasibility, accuracy, and efficiency.

International Journal of Advance Interdisciplinary Research
Vol. 2, Special Issue 1, (Jan-March) 2026, pp. 33-45, e-ISSN: 3107-913X
DOI: https://doi.org/10.66095/ijair.2026.v2.s1.4

Data and Methodology

- **Data Collection:** Using simulated and real-world datasets, this study investigates various probability designs, such as simple random sampling, stratified random sampling, systematic sampling cluster sampling and multi-stage sampling for their relevance and efficiency. To evaluate the efficacy of different sampling techniques, important factors like population size, size of sample, and secondary variables are considered.
- **Methodology:** Simple Random Sampling, Stratified Sampling, Systematic Selection, Cluster-Based Sampling, and Multi-Level Sampling. are among the probability sampling techniques that are used. Through comparative analysis and simulation studies, their efficiency is assessed while taking statistical metrics like precision, mean square error (MSE), and bias into account.

Existing estimators

Cochran (1940) proposed the standard ratio estimator, which is provided by:

$$t_r = \frac{\bar{X}}{x}\bar{y}$$

Bias (t_r)= **0**

MSE (t_r)= $f\bar{Y}^2(C_y^2 + C_x^2 - 2\rho_{xy}C_xC_y)$

Watson (1937) proposed the standard regression estimator, which is provided by

$$t_{reg} = \bar{y} + \beta^*(\bar{X} - \bar{x})$$

with the minimum MSE and bias at the ideal scalar value

$$\beta = \rho_{xy}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}$$

Bias (t_{reg})= $f\bar{Y}\theta(\theta C_x^2 - \rho_{xy}C_xC_y)$

MSE (t_{reg}) = $f\bar{Y}^2 Cy^2(1-\rho xy^2)$

Srivastava (1967) proposed the following power ratio estimator.

$$t_s = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\delta}$$

International Journal of Advance Interdisciplinary Research
Vol. 2, Special Issue 1, (Jan-March) 2026, pp. 33-45, e-ISSN: 3107-913X
DOI: https://doi.org/10.66095/ijair.2026.v2.s1.4

with the minimum MSE and bias at the ideal scalar value

$$\delta = \rho_{xy}\frac{C_y}{C_x}$$

Bias (t_s)= $f\bar{Y}\delta(\left(\frac{\delta+1}{2}\right)C_x^2 - \rho_{xy}C_xC_y)$

MSE (t_s)= f$\bar{Y}^2$Cy^2(1−ρxy^2)

Walsh (1970) proposed the ratio estimator that follows:

$$t_w = \bar{y}(\frac{\bar{x}}{X + \theta\overline{(x - \bar{X})}})$$

with the minimum MSE and bias at the ideal scalar value

$$\theta = \rho_{xy}\frac{C_y}{C_x}$$

Bias (t_w)= $f\bar{Y}\theta(\theta C_x^2 - \rho_{xy}C_xC_y)$

MSE (t_w)= f$\bar{Y}^2$Cy^2(1−ρxy^2)

Sisodia and Dwivedi (1981) utilized the auxiliary information and proposed the following ratio estimator

$$t_{Sd} = \bar{y}\left(\frac{\bar{X} + C_x}{\bar{x} + C_x}\right)$$

Bias (t_{sd})=$f\bar{Y}\lambda_1\ (\lambda_1 C_x^2 - \rho_{xy}C_xC_y)$

MSE (t_{sd})= $f\bar{Y}^2(C_y^2 + \lambda_1^2 C_x^2 - 2\lambda_1\rho_{xy}C_xC_y)$

$$\lambda_1 = \frac{\bar{X}}{\bar{X} + C_x}$$

Bahl and Tuteja (1991) suggested the exponential ratio estimator as

t_{re}=$\bar{y}\ exp(\frac{\bar{X}-\bar{x})}{\bar{X}+\bar{x}})$

Bias (t_{re})=$f\bar{Y}(\frac{3}{8}C_x^2 - \frac{1}{2}\rho_{xy}C_xC_y)$

MSE (t_{re})= $f\bar{Y}^2(C_y^2 + \frac{C_x^2}{4} - \rho_{xy}C_xC_y)$

Singh and Kakran (1993) used the coefficient of kurtosis and establish the following ratio estimator

$$t_{sk} = \bar{y}\left(\frac{\bar{X} + \beta_2(x)}{\bar{\bar{X}} + \beta_2(x)}\right)$$

Bias (t_{sk})=$f\bar{Y}\lambda_2(\lambda_2 C_x^2 - \rho_{xy}C_x C_y)$

MSE (t_{sk})= $f\bar{Y}^2(C_y^2 + \lambda_2^2 C_x^2 - 2\lambda_2\rho_{xy}C_x C_y)$

$$\lambda_2 = \frac{\bar{X}}{\bar{X} + \beta_2(x)}$$

Yadav and Kadilar (2013) proposed an enhanced category of estimators for exponential ratios:

$$t_{yk} = k\bar{y}\, exp\left\{\frac{(U\bar{X} + v) - (\overline{Ux} - v)}{(U\bar{X} + v) + (\overline{Ux} - v)}\right\}$$

Bias (t_{yk})= $kf\bar{Y}\left(2\tau^2 C_x^2 - \tau\rho_{xy}C_x C_y\right) + \bar{Y}(k+1)$

MSE (t_{yk})= $\bar{Y}^2\left(1 - \frac{A^2}{B}\right)$

A specific class of estimators was proposed by **Ijaz and Ali (2018)** for estimation of population mean.

$$t_{ia} = W\bar{y} + (1 - W)\bar{y}\frac{\bar{X}}{\bar{x}}$$

Bias (t_{ia})= $(1 - w)f\bar{Y}(C_x^2 - \rho_{xy}C_x C_y)$

MSE (t_{ia})= $f\bar{y}^2 Cy^2(1-\rho xy^2)$

Singh et al. (2009) proposed the following group of exponential estimators after considering a variety of auxiliary data:

$$t_{se} = \bar{y}\, exp\left\{\frac{(U\bar{X}+v)-(\overline{Ux}-v)}{(U\bar{X}+v)+(\overline{Ux}-v)}\right\}$$

Bias (t_{se})= $f\bar{Y}(\tau^2 C_x^2 - \tau\rho_{xy}C_x C_y)$

MSE (t_{se})= $f\bar{Y}^2(C_y^2 + \tau^2 C_x^2 - 2\tau\rho_{xy}C_x C_y)$

$$\tau = \frac{a\bar{X}}{2(a\bar{X}+b)}$$

An enhanced ratio cum exponential ratio class of estimator was created by **Kadilar (2016):**

$t_{gk}=\bar{y}\left(\frac{\bar{x}}{\bar{x}}\right)^{\alpha} exp(\frac{\bar{X}-\bar{x})}{\bar{X}+\bar{x}})$

Bias $(t_{gk})=f\bar{Y}[\frac{\alpha(1+\alpha)}{2}C_x^2+\frac{\alpha}{2}C_x^2+\frac{3}{8}C_x^2-\left(\alpha+\frac{1}{2}\right)\rho_{xy}C_xC_y]$

MSE $(t_{gk})=f\bar{Y}^2C_y^2(1-\rho_{xy}^2)$

Proposal Estimator

A proposal for an estimator is a statistical method that uses auxiliary data to improve population parameter estimation's precision and dependability. By reducing bias and the mean square error (MSE) through adjustments based on known population characteristics, it increases efficiency. By adding auxiliary variables like mean, variance, and coefficients of correlation to improve accuracy, this estimator improves conventional techniques. It improves estimation accuracy by incorporating more data, especially in situations where traditional probability sampling methods might introduce errors. The proposal estimator, which is frequently employed in research surveys, econometrics, and social sciences, guarantees more accurate and reliable analysis of statistics and inference.

$$t_{new}=\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\delta}\exp\left\{\frac{\theta\overline{(X-x)}}{\bar{X}+\theta\bar{x}}\right\}$$

Bias $(t_{new})=\bar{Y}[\left(\frac{1}{n}-\frac{1}{N}C_x^2\right)\left(\frac{3}{2}\gamma^2+\delta\gamma+\frac{\delta(\delta+1)}{2}\right)-\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}C_yC_x(\gamma+\delta)]$

MSE $(t_{new})=\bar{y}\ [(\gamma+\delta)^2\left(\frac{1}{n}-\frac{1}{N}\right)C_x^2+\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2+\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}C_yC_x(\gamma+\delta)]$

$$\gamma=\frac{\theta}{1+\theta}$$

$$\delta=-2,-1,0,1,2$$

$$\theta=C_x^2,C_y^2,\rho_{xy}$$

Efficiency Comparison

International Journal of Advance Interdisciplinary Research
Vol. 2, Special Issue 1, (Jan-March) 2026, pp. 33-45, e-ISSN: 3107-913X
DOI: https://doi.org/10.66095/ijair.2026.v2.s1.4

$$(t_{reg}, t_{new}) = \frac{f\bar{Y}^2 C_y^2(1-\rho_{xy}^2)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_s, t_{new}) = \frac{f\bar{Y}^2 C_y^2(1-\rho_{xy}^2)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_w, t_{new}) = \frac{f\bar{Y}^2 C_y^2(1-\rho_{xy}^2)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{ia}, t_{new}) = \frac{f\bar{Y}^2 C_y^2(1-\rho_{xy}^2)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_r, t_{new}) = \frac{f\bar{Y}^2((C_y^2 + C_x^2 - 2\rho_{xy} C_x C_y)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{re}, t_{new}) = f\bar{Y}^2(C_y^2 + \frac{C_x^2}{4} - \frac{2\rho_{xy} C_x C_y}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{sd}, t_{new}) = \frac{f\bar{Y}^2(C_y^2 + \lambda_1^2 C_x^2 - 2\lambda_1 \rho_{xy} C_x C_y)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{sk}, t_{new}) = \frac{f\bar{Y}^2(C_y^2 + \lambda_2^2 C_x^2 - 2\lambda_2 \rho_{xy} C_x C_y)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{s1}, t_{new}) = \frac{f\bar{Y}^2(C_y^2 + \lambda_5^2 C_x^2 - 2\lambda_5 \rho_{xy} C_x C_y)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{se}, t_{new}) = \frac{f\bar{Y}^2(C_y^2 + \tau^2 C_x^2 - 2\tau \rho_{xy} C_x C_y)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(\hat{y}_R^*, t_{new}) = 1 - \frac{f\bar{Y}^2[C_y^2 + C_x^2 \lambda_i(\lambda_i - 2\theta)]}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

$$(t_{yk}, t_{new}) = \frac{\bar{Y}^2\left(1 - \frac{A^2}{B}\right)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

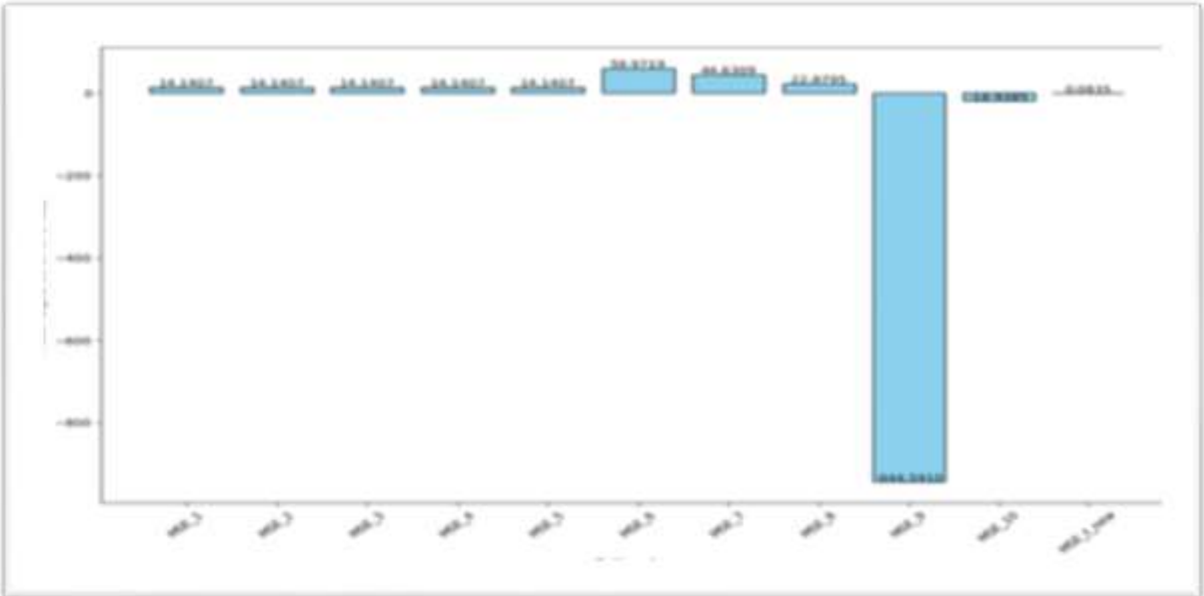
$$(t_{gk}, t_{new}) = \frac{f\bar{Y}^2 C_y^2(1-\rho_{xy}^2)}{[(\gamma+\delta)^2 C_x^2 + C_y^2 + \rho_{xy} C_y C_x(\gamma+\delta)]} * 100$$

Table 1: Comparison of Population 1, 2 and 3

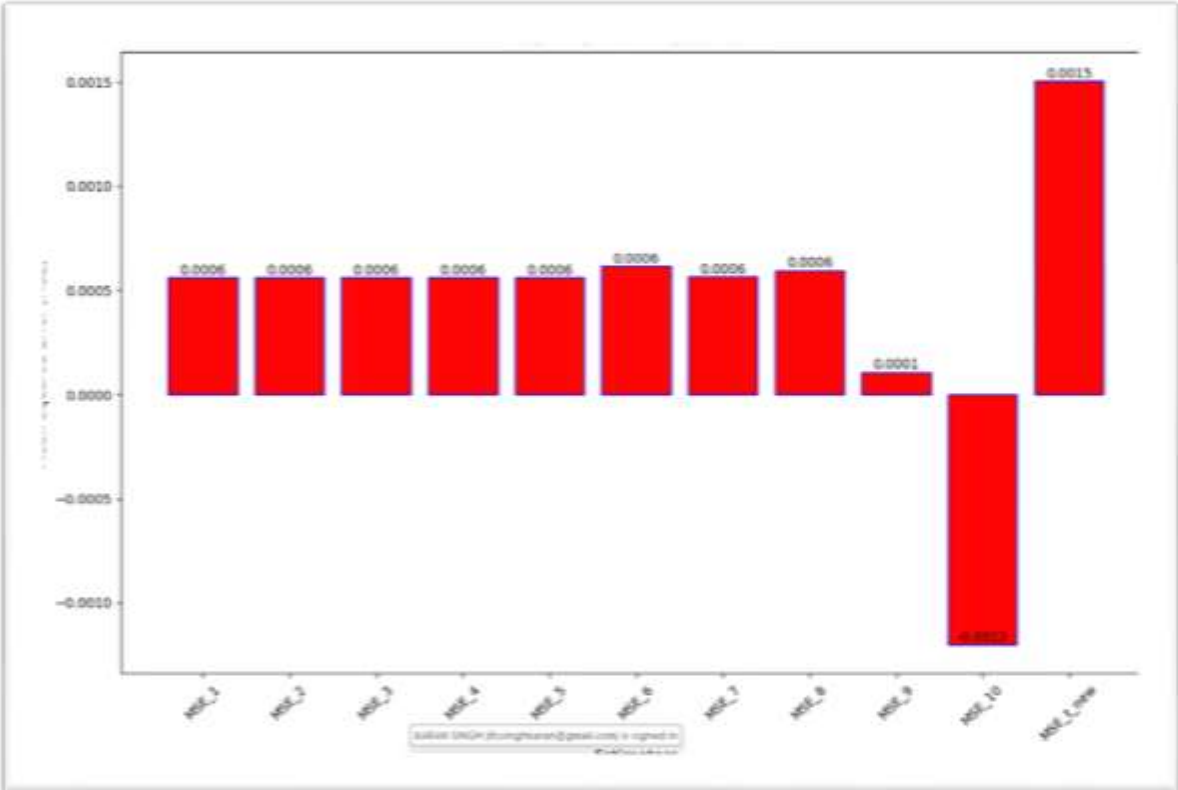
Parameters	Population 1: Source: Sarndal et al. (2003, 652–659), Y = Total	Population 2: Source: Singh (2003b, 1115), Y =	Population 3: Source: Alomair and Shahzad (2023), Y =

	number of seats (S82) in municipal council in 1982 and X = Number of conservative seats (CS82) in municipal council in 1982.	Seasonal average price per pound during 1996 and X = Seasonal average price per pound during 1995.	Humidity (%) in Karachi during 2022 and X = Humidity (%) in Karachi during 2021.
N	284	36	365
n	50	14	100
$\bar{Y}$	46.07	0.20	90.09
$\bar{X}$	9.09	0.18	90.82
Mx	8.00	0.18	91
Cy	0.27	0.39	0.08
Cx	0.54	0.40	0.08
ρxy	0.68	0.87	0.75
$\beta 1(x)$	1.24	0.77	−0.48
$\beta 2(x)$	5.38	3.34	3.21

Results have shown in Bar graph



[Figure 1]



[Figure 2]

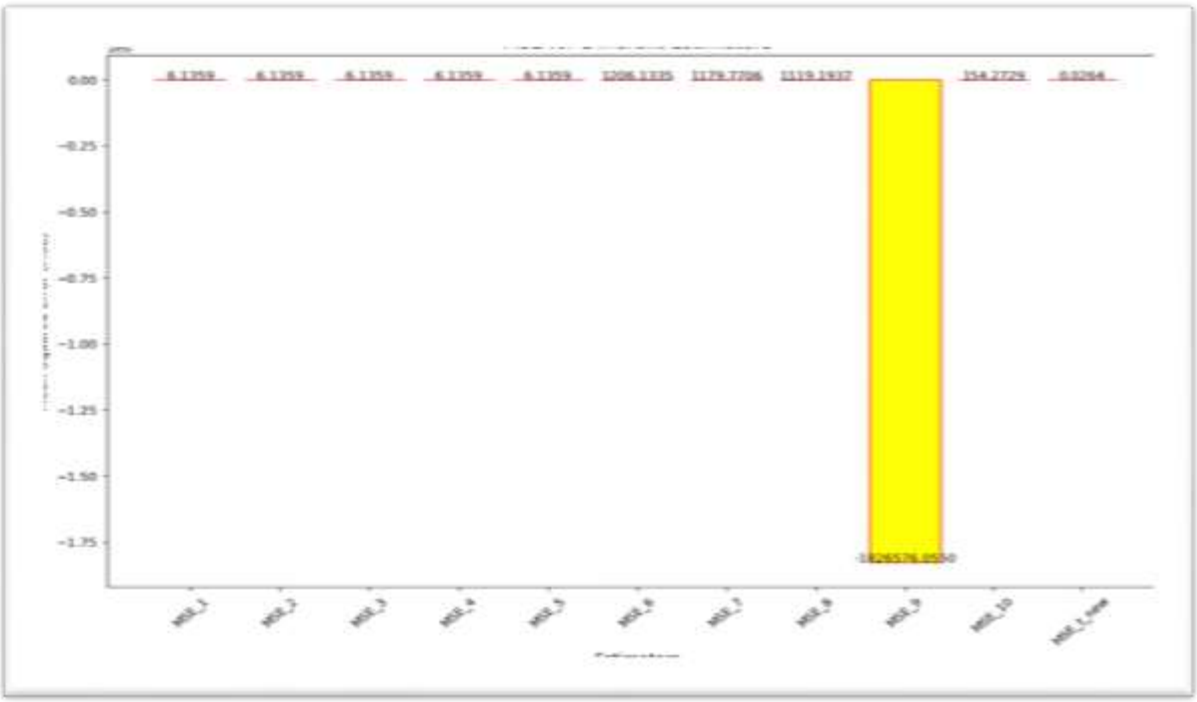


Figure [3]

Simulation Study

To extend the mathematical results and support the outcomes of the numerical analysis, we performed a study using simulations using either symmetric and asymmetric data [6]. In accordance with Singh and Horn's (1998) methodology, we generated the data using the models listed below:

$$y = 3.4 + \sqrt{\left(1 - \rho_{x_y}\right)}\ y^* + \rho_{xy}\left(\frac{S_x}{S_y}\right)x^*$$

$$x = 3.2 + x^*$$

(1) We generated a normal population of size N=10000using x∗ ~ N (15, 65) and y∗ ~ N (20, 70).

(2) We have done simulation on real population on data Sarndal et al. (2003,652-659), Y=Total number of seats(S82) in municipal council in 1982 and X=Number of conservative seats (CS82) in municipal council in 1982.

Percent Relative Efficiency (PRE)

The performance of the suggested estimators was assessed through a simulation study. To ensure a representative portion, a collection of n = 5000was chosen from an actual population of N = 1000. No of samples were selected to guarantee consistency and reduce sampling variability [1]. 10,000 replicas of the sampling procedure were used in the study, enabling precise performance evaluation. For each estimator, the Mean Square of Error (MSE) and The Percentage Relative Efficiency (PRE) were computed. This made it easier to compare their efficiency and accuracy. A standard formula was used to calculate the PRE, which shed light on each estimator's dependability in statistical analysis.

$$PRE = \frac{\sum_{i=1}^{10000}(t_m - \bar{Y})^2}{\sum_{i=1}^{10000}(T^* - \bar{Y})^2} \times 100$$

For first population (1)

In addition to the proposal estimator, estimates from Prasad (1989), Srivastava (1967), Kadilar and Cingi (2006), and Singh and Talior (2003) have been used for the simulation study

At $\delta = -2, \theta = 1$

To evaluate how well different estimators perform for a simulate normal population Bias, MSE, PRE.

Table 2

parameters	Bias	MSE	PRE
t_1	0.000	0.017	30542.7
t_2	0.000	0.021	25278.2
t_3	0.000	0.017	30552.2
t_4	-0.49	24.3	0.222
t_5	-0.000	0.038	14299.2
t_p	-0.000	0.022	24849.3

At $\delta = 2, \theta = 0$

To evaluate how well different estimators perform for a simulate normal population Bias, MSE, PRE.

Table 3

parameters	Bias	MSE	PRE
t_1	0.000	0.028	19461
t_2	0.000	0.024	22196
t_3	0.000	0.028	19465
t_4	-0.49	24.3	0.222

t_5	-0.000	0.019	27722
t_p	-0.000	0.022	24849

At $\delta = 1, \theta = 2$

To evaluate how well different estimators perform for a simulate normal population Bias, MSE, PRE

Table 4

parameters	Bias	MSE	PRE
t_1	-0.0005	0.016	33911
t_2	-0.0005	0.013	41561
t_3	-0.0005	0.016	33913
t_4	-0.49	24.4	0.223
t_5	-0.0005	0.016	33911
t_p	-0.0006	0.019	27886

At $\delta = 0, \theta = 1$

To evaluate how well different estimators perform for a simulate normal population Bias, MSE, PRE

Table 5

parameters	Bias	MSE	PRE
t_1	0.000	0.028	19461

International Journal of Advance Interdisciplinary Research
Vol. 2, Special Issue 1, (Jan-March) 2026, pp. 33-45, e-ISSN: 3107-913X
DOI: https://doi.org/10.66095/ijair.2026.v2.s1.4

t_2	0.0006	0.024	22196
t_3	0.0006	0.028	19465
t_4	-0.49	24.5	0.2247
t_5	0.000	0.019	27722
t_p	0.000	0.022	24849

Simulation on real population (2)

In addition to the proposal estimator, estimates from Prasad (1989), Srivastava (1967), Kadilar and Cingi (2006), and Singh and Talior (2003) have been used for the simulation study.

Table 6

parameters	Bias	MSE	PRE
t_1	-0.08	4.62	118.5
t_2	-0.06	4.58	119.6
t_3	-0.08	3.56	153.8
t_4	-42.3	1789.8	0.306
t_5	0.10	0.76	715.2
t_p	-0.02	0.22	2407

At $\delta = 0, \theta = 1$

Table 7

parameters	Bias	MSE	PRE
t_1	0.07	15.33	35.7
t_2	0.10	15.41	35.5
t_3	0.04	10.9	49.8
t_4	-42.3	1785.6	0.306
t_5	0.07	15.33	35.7
t_p	0.07	15.33	35.7

At $\delta = 1, \theta = 0$

Table 8

Table 9

International Journal of Advance Interdisciplinary Research
Vol. 2, Special Issue 1, (Jan-March) 2026, pp. 33-45, e-ISSN: 3107-913X
DOI: https://doi.org/10.66095/ijair.2026.v2.s1.4

parameters	Bias	MSE	PRE
t_1	0.25	5.58	94.02
t_2	0.28	5.75	95.15
t_3	0.19	4.29	127.57
t_4	-42.3	1793.3	0.305
t_5	0.25	5.82	94.02
t_p	0.63	8.63	63.48

At $\delta = 1, \theta = 2$

parameters	Bias	MSE	PRE
t_1	0.14	10.6	51.47
t_2	0.16	10.5	51.91
t_3	0.09	7.7	71.08
t_4	-42.3	1791.6	0.305
t_5	0.09	48.0	11.4
t_p	0.02	11.2	48.6

At $\delta = -1, \theta = 2$

Result & Discussion

The study assesses probability sampling methods according to accuracy, bias reduction, and mean squared error (MSE). The simple random sampling method (SRS) has a high variance even though it provides unbiased selection. The most accurate technique is stratified sampling since it successfully lowers variability. For structured populations, systematic sampling works well, but if there are recurring patterns, bias may be introduced. Large-scale studies benefit from cluster and multi-stage sampling; however, because of similarities within clusters, their MSE is typically higher. These results emphasize how crucial it is to choose a sampling strategy that is appropriate for data analysis to give accurate and efficient results.

References

1. Anoop Kumar & Asra Sayyed Siddiqui (2024) Enhanced estimation of population mean using simple random sampling, Research in Statistics, 2:1, 2335949, DOI: 10.1080/27684520.2024.2335949
2. Bahl, S., & Tuteja, R. (1991). Exponential ratio estimators in simple random sampling.
3. Cochran, W. G. (1940). The use of analysis of variance in enumeration by sampling.
4. Singh, H. P., Tailor, R., Das, A., & Chowdhury, J. (2009). A generalized exponential ratio-type estimator for finite population means.

5. Watson DJ. 1937. The estimation of leaf area in field crops. J Agric Sci. 27:474–483.

6. Watson, G. S. (1937). Ratio and regression estimation in sample surveys.