

RESEARCH

Open Access



A robust population mean estimation method in stratified sampling using auxiliary attributes under real and simulated populations

Mukesh Kumar Verma¹ and Saif Ali Khan^{2*}

*Correspondence:
saifali.khan@manipal.edu

¹ Department of Community
Medicine, Hind Institute
of Medical Sciences, Sitapur,
India

² Directorate of Online Education,
Manipal Academy of Higher
Education, Manipal, India

Abstract

This study proposes a novel improved ratio-type estimator for efficient estimation of the population mean in stratified sampling by integrating auxiliary attribute information through a unified nonlinear framework. Unlike conventional ratio, product, and exponential-type estimators, the proposed estimator combines these estimation strategies within a single flexible formulation, allowing it to adapt to different correlation structures and improve estimation efficiency. The mean squared error (MSE) of the proposed estimator is derived up to the first order of approximation, and the optimum parameter values that minimize the MSE are obtained analytically. The estimator's performance is evaluated using percentage relative efficiency (PRE) and compared with existing estimators reported in the literature. Numerical illustrations and extensive Monte Carlo simulation studies demonstrate consistent efficiency gains across diverse sampling scenarios. The practical applicability of the proposed estimator is further validated using a real-world apple production dataset, where apple yield is treated as the study variable and the number of apple trees as the auxiliary attribute. Both theoretical and empirical results confirm that the proposed estimator provides more precise and robust population mean estimates than existing approaches under stratified sampling.

Keywords: Auxiliary variable, Auxiliary attribute, Robust approach, MSE, PRE

Introduction

Accurate estimation of population parameters is a primary objective in survey sampling, particularly for heterogeneous populations. Stratified sampling is widely used because it improves the precision of estimates by dividing the population into relatively homogeneous strata. The efficiency of estimators can be further enhanced by incorporating auxiliary information that is strongly associated with the study variable. Conventional ratio, product, and exponential estimators have been widely applied for this purpose, although their performance depends on the strength and nature of the underlying relationship. In agricultural surveys, such as apple production studies, auxiliary attributes are often readily available and closely related to crop yield. Proper use of these attributes can

substantially improve estimation accuracy without increasing survey costs. Despite their practical importance, auxiliary attributes remain underutilized in stratified sampling.

Motivated by this gap, the present study proposes a new ratio-type estimator that incorporates auxiliary attributes under stratified sampling. Theoretical expressions for its mean squared error are derived using first-order approximation, and its performance is evaluated through percent relative efficiency. The proposed estimator is shown to provide more accurate and efficient estimates than existing methods, demonstrating its practical value for agricultural and other survey-based applications.

The stratification concept, which is applied to improve the accuracy of the study of the population, could be used to divide it into two classes according to an attribute. Therefore, the units of the population have been divided into these two types according to whether or not they possess a particular attribute. Once we have a sample of size n , the population proportion of the defined property may prove useful for our objectives. In order to enhance the field of survey sampling, Kaur [1] investigates the population mean estimate in the stratified sampling scheme. The exponential estimator has been introduced by Bahl and Tuteja [2] in simple random sampling (SRS). In stratified random sampling, in order to improve the accuracy and efficacy of population mean estimates, Diana [3] introduces and examines a variety of new estimators. In order to enhance the estimation of population parameters in stratified random sampling, Kadilar and Cingi [4] evaluate ratio estimators for the purpose of to improve population estimates' efficiency and accuracy, the study develops new or modified ratio estimators and utilizes the use of auxiliary variables. Ratio and product estimators have been discussed and evaluated by Koyuncu and Kadilar [5] with the aim of enhancing the stratified random sampling technique's population parameter estimation. The study aims to improve the efficiency and accuracy of population estimates by utilizing auxiliary variables and developing new or improved ratio and product estimators. Enhanced multivariate-ratio-type estimators are presented and examined by Malik and Singh [6] in an effort to increase the population estimate parameters while using stratified random sampling. Lu [7] proposes and examines an effective estimator that combines two auxiliary variables to enhance population mean estimate. Zaman [8] proposes and evaluates new estimators which utilize auxiliary attributes to enhance population mean estimate. Zaman and Kadilar [9] discuss and evaluate novel techniques for utilizing auxiliary characters to estimate population means in an approach to stratified random sampling. Akintunde et al. [10] proposes novel techniques for utilizing auxiliary variables used to estimate the population mean in a situation of stratified sampling. In survey sampling and statistical estimation, an auxiliary attribute is a variable or characteristic that is not the primary variable of interest but is connected to it in a specific way. Zaagan et al. [11] explored the use of auxiliary attributes to enhance precision, demonstrating through analytical and numerical results that their estimators outperformed conventional stratified methods. Daraz et al. [12] developed improved estimators for the finite population median and variance using auxiliary variables, Daraz et al. [13] developed improved Monte Carlo simulations, and real datasets, demonstrating superior performance over conventional methods and also Daraz et al. [14] proposed efficient median estimators for stratified multi-population data, reporting substantial gains in estimation

accuracy for health-related applications. Verma et al. [15] introduced an uncertainty-based framework that explicitly accounted for variability in the estimation process, leading to improved efficiency and reduced mean squared error, as supported by simulation and numerical studies. Addressing practical constraints, Verma et al. [16] incorporated a linear cost function into the estimation process and showed that optimal allocation across strata resulted in greater efficiency under budget limitations, also Verma [17] extended this line of work to the joint estimation of two population means, proposing optimal estimators that leveraged auxiliary information to achieve lower mean squared error than existing approaches. Yadav et al. [18] demonstrate the efficiency of a suggested estimator over conventional methods by obtaining bias and mean squared error equations. Chaudhary et al. [19] developed calibration-based estimators for stratified sampling that handled both non-response and measurement errors. Using auxiliary information and optimized weights, the proposed estimators reduced bias and MSE and performed better than existing methods according to theoretical and empirical comparisons. Kumar et al. [20] developed direct and synthetic logarithmic estimators for small area estimation using ranked set sampling, which improved efficiency and provided more accurate estimates compared to existing methods. Robbennolt et al. [21] examined data collection, weighting, and modeling techniques to estimate consistent population parameters, highlighting methods that improved accuracy and ensured reliable inference in transportation studies.

The studies collectively focused on advancing population mean estimation under different sampling frameworks. The Verma et al. [22] study developed an optimal strategy in stratified random sampling using an auxiliary attribute to improve efficiency over traditional estimators. The Verma et al. [23] work introduced neutrosophic approaches that accounted for uncertainty and applied them to Indian temperature data. In Verma et al. [24], optimal estimation techniques were explored within stratified sampling with practical application in supply chain management for improved decision-making. Finally, the Verma et al. [25] study proposed an improved estimator in post-stratified sampling, supported by applications and simulation-based evaluation, demonstrating enhanced performance.

The proposed estimator is developed under the first-order approximation, where higher-order terms are ignored. As a result, the derived bias and mean squared error (MSE) expressions are approximate and may be less accurate for small sample sizes or complex population structures. The methodology also assumes that the auxiliary attribute is measured without error and is strongly associated with the study variable. Moreover, the empirical evaluation is based on a limited dataset. Future research may consider second-order approximations, measurement errors, non-response, and broader simulation studies to further assess the robustness of the proposed estimator.

The proposed estimator was evaluated using both simulated and real datasets, including an apple production survey where apple yield was the study variable and the number of apple trees served as the auxiliary attribute. The results consistently demonstrated superior performance over existing estimators, confirming its efficiency and robustness. These findings highlight the effectiveness of incorporating auxiliary attributes in stratified sampling and support the applicability of the proposed estimator across agriculture, health, environmental, and socio-economic surveys.

Notations and formulation

Let us consider a size of random sample is n and a finite population of size n_h is drawn using SRSWOR from the h th stratum such that it is divided into L strata. Let ϕ is an auxiliary attribute and Y be the study variable and taking values ϕ_{hi} and y_{hi} respectively ($h = 1, 2, \dots, L; i = 1, 2, \dots, N_h$) on i th unit of the h th stratum. Each stratum provides a sample of size n_h to obtain a sample of size n . With $i = 1, 2, \dots, N_h$ and $h = 1, 2, \dots, L$ let y_{hi} and ϕ_{hi} represent the study variable and auxiliary attribute ϕ respectively. Assume further that either the presence or missing an attribute ϕ_h implies a complete dichotomy in the population and that there are actually only two possible are 1 and 0 (1 is i th unit of population possesses attribute ϕ and 0 take other) values for the attribute ϕ_h . Let $A = \sum_{i=1}^N \phi_i$, $A_h = \sum_{i=1}^{N_h} \phi_{hi}$, $a = \sum_{i=1}^n \phi_i$ and $a_h = \sum_{i=1}^{n_h} \phi_{hi}$ is the total number of units with attribute ϕ in the population mean and sample mean in h th stratum respectively. Let $P = \frac{A}{N}$, $P_h = \frac{A_h}{N_h}$, $p = \frac{a}{n}$ and $p_h = \frac{a_h}{n_h}$. The followings may be considered as, $\bar{Y}_h$ is mean for the study variable and $\bar{\phi}_h$ is mean for the auxiliary attribute. The population mean of study variable is $\bar{Y} = \sum_{h=1}^L W_h \bar{Y}_h$ and $W_h = \frac{N_h}{N}$ is weight of h th stratum. Let $\bar{y}_{st} = \bar{Y}(1 + e_0)$ and $p_{st} = P(1 + e_1)$, where e_0 and e_1 are errors of approximation and the value expectations of e_0 and e_1 , we have $E(e_0) = 0$ and $E(e_1) = 0$. Let, $E(e_0^2) = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2} = \omega_{2.0}$, $E(e_1^2) = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{P^2} = \omega_{0.2}$ and $E(e_0 e_1) = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{P \bar{Y}} = \omega_{1.1}$. $\omega_{a.b} = \sum_{h=1}^L W_h^{a+b} \lambda_h \frac{E(y_{hi} - \bar{Y}_h)^a (p - P)^b}{\bar{Y}^a P^b}$ and $\lambda_h = \left(\frac{1}{n_h} - \frac{1}{N_h} \right)$; $h = 1, 2, \dots, L$. S_{yh}^2 , S_{ph}^2 and S_{yph} represents the population variances of the study variable and auxiliary attribute, and the covariance between study variable and auxiliary attribute the population in the h th stratum respectively. First-order approximations are used because they provide simple, tractable expressions for bias and mean squared error, making analysis and comparison of estimators easier. They are sufficiently accurate for moderate to large samples and are standard in sampling theory, while higher-order terms add complexity with limited practical gain.

Some existing estimators

In stratified random sampling, the sample mean is

$$\bar{y}_{st} = t_{1(st)} = \sum_{h=1}^L W_h \bar{y}_h \quad (1)$$

and variance is

$$Var(\bar{y}_{st}) = MSE(t_{1(st)}) = \sum_{h=1}^L W_h^2 \lambda_h S_{yh}^2 = \bar{Y}^2 \omega_{2.0} \quad (2)$$

Naik and Gupta [26] proposed the ratio estimator in auxiliary attribute is as follow

$$t_{2(st)} = \bar{y}_{st} \left[\frac{P}{P_{st}} \right] \quad (3)$$

and its MSE is

$$MSE(t_{2(st)}) = \bar{Y}^2 [\omega_{2.0} - 2\omega_{1.1} + \omega_{0.2}] \quad (4)$$

where $\omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}$, $\omega_{0.2} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{\bar{p}^2}$ & $\omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{\bar{p}\bar{Y}}$.

The exponential ratio type estimator was presented by Bahl and Tuteja [2] and is as follows in the case of an auxiliary attribute.

$$t_{3(st)} = \bar{y}_{st} \exp \left[\frac{P - p_{st}}{P + p_{st}} \right] \quad (5)$$

and MSE is as follow

$$MSE(t_{3(st)}) = \bar{Y}^2 \left[\omega_{2.0} - \omega_{1.1} + \frac{1}{4} \omega_{0.2} \right] \quad (6)$$

where $\omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}$, $\omega_{0.2} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{\bar{p}^2}$ & $\omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{\bar{p}\bar{Y}}$.

Zaman [8] examine the effective population mean estimators utilizing auxiliary variables in stratified random sampling:

$$t_{4(st)} = \bar{y}_{st} \frac{\sum_{h=1}^L W_h (P_h + C_{ph})}{\sum_{h=1}^L W_h (p_{st} + C_{ph})} \quad (7)$$

and MSE is

$$MSE(t_{4(st)}) = \bar{Y}^2 \left[\omega_{2.0} - 2\delta_1 \omega_{1.1} + \delta_1^2 \omega_{0.2} \right] \quad (8)$$

where $\delta_1 = \frac{\sum_{h=1}^L W_h P_h}{\sum_{h=1}^L W_h (P_h + C_{ph})}$, $\omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}$, $\omega_{0.2} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{\bar{p}^2}$ & $\omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{\bar{p}\bar{Y}}$.

Zaman and Kadilar [9] examine improved general category of exponential ratio type estimators with an auxiliary attributes is as follow

$$t_{5,i(st)} = k \bar{y}_{st} \exp \left[\frac{(kp_{st} + l) - (kp_{st} + l)}{(kp_{st} + l) + (kp_{st} + l)} \right] \quad (9)$$

and MSE is

$$MSE(t_{5,i(st)}) = \bar{Y}^2 \left[k^2 \omega_{2.0} + Q_i^2 (3k^2 - 2k) \omega_{0.2} - 2Q_i (3k^2 - 2k) \omega_{1.1} + (k - 2)^2 \right] \quad (10)$$

where $k = \frac{P}{P+1}$ is an appropriate constant that will be chosen later. Here

$$Q_1 = \frac{kP}{2(kP+1)}, \quad Q_2 = \frac{P}{2(P+\beta_2)}, \quad Q_3 = \frac{P}{2(P+C_p)}, \quad Q_4 = \frac{P}{2(P+\rho_p)}, \quad Q_5 = \frac{\beta_2 P}{2(\beta_2 P + C_p)},$$

$$Q_6 = \frac{C_p P}{2(\beta_2 + C_p P)}, \quad Q_7 = \frac{\beta_2 P}{2(\rho_p + C_p P)}, \quad Q_8 = \frac{\rho_p P}{2(\rho_p P + C_p)}, \quad Q_9 = \frac{\beta_2 P}{2(\beta_2 P + \rho_p)}, \quad Q_{10} = \frac{\rho_p P}{2(\rho_p P + \beta_2)}$$

$$k = \frac{2Q_i^2 \omega_{0.2} - 2Q_i \omega_{1.1} + 2}{6Q_i^2 \omega_{0.2} - 8Q_i \omega_{1.1} + 2\omega_{2.0} + 2}, \text{ for all } i = 1, 2, \dots, 10.$$

$$\omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}, \omega_{0.2} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{\bar{p}^2} \text{ \& } \omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{\bar{p}\bar{Y}}.$$

Akintunde et al. [10] examine the enhanced population mean estimation in stratified sampling with an auxiliary attributes is as follows.

$$t_{6(st)} = \bar{y}_{st} \left[\alpha \exp \left(\frac{(\varphi P_{st} + L) - (\varphi P_{st} + L)}{(\varphi P_{st} + L) + (\varphi P_{st} + L)} \right) + (1 - \alpha) \exp \left(\frac{(\varphi P_{st} + L) - (\varphi P_{st} + L)}{(\varphi P_{st} + L) + (\varphi P_{st} + L)} \right) \right] \quad (11)$$

and MSE is

$$MSE(t_{6(st)}) = \bar{Y}^2 \left[\omega_{2.0} + \zeta_1^2 (4\alpha^2 - 4\alpha + 1) \omega_{0.2} - 2\zeta_1 (2\alpha - 1) \omega_{1.1} \right] \quad (12)$$

where, $\zeta_1 = \frac{\varphi P}{2(\varphi P + 1)}$ and $\alpha = \frac{\omega_{1.1}}{2\omega_{0.2}\zeta_1} + \frac{1}{2}$ is constant value.

$$\omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}, \omega_{0.2} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{P^2} \text{ \& } \omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{P\bar{Y}}.$$

Verma et al. [22] estimated the population mean in stratified random sampling using auxiliary attribute information is

$$t_{7(st)} = \bar{y}_{st} \left[\frac{P}{p_{st}} \right] + z_1 \bar{y}_{st} \exp \left[\frac{P - p_{st}}{P + p_{st}} \right] + z_2 \bar{y}_{st} \left[1 + \log \left(\frac{p_{st}}{P} \right) \right] \quad (13)$$

with MSE is,

$$MSE(t_{7(st)}) = \bar{Y}^2 \left[a + z_1^2 b + z_2^2 c + 2z_1 d + 2z_2 p + 2z_1 z_2 q \right] \quad (14)$$

where $a = \omega_{2.0} + \omega_{0.2} - 2\omega_{1.1}$, $b = 1 + \omega_{2.0} + \omega_{0.2} - 2\omega_{1.1}$, $c = 1 + \omega_{2.0} + 4\omega_{1.1}$, $d = \omega_{2.0} - \frac{5}{2}\omega_{1.1} + \frac{3}{2}\omega_{0.2}$, $p = \omega_{2.0} - \omega_{1.1}$, $q = 1 + \omega_{2.0} + \omega_{1.1} - \frac{3}{8}\omega_{0.2}$, $z_1 = \frac{cd - pq}{q^2 - bc}$ and $z_2 = \frac{bp - dq}{q^2 - bc}$.

$$\omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}, \omega_{0.2} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{P^2} \text{ \& } \omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{P\bar{Y}}.$$

Proposed estimator

Motivated by Akintunde et al. [10], we propose an enhanced ratio-type estimator under stratified random sampling by effectively incorporating auxiliary attribute information within a modified functional framework. The mean squared error (MSE) of the estimator is derived up to the first order of approximation. The proposed estimator is presented as follows:

$$t_{pr(st)} = \bar{y}_{st} + K_1 \bar{y}_{st} \left[\frac{P}{p_{st}} \right] + K_2 \bar{y}_{st} \exp \left[\frac{P - p_{st}}{P + p_{st}} \right] \quad (15)$$

where K_1 and K_2 are optimum constant, put value of $\bar{y}_{st}$ and p_{st} in above equation and simplify

$$t_{pr(st)} = \bar{Y} \left[(1 + e_0) + K_1 \left(1 + e_0 - e_1 - e_0 e_1 + e_1^2 \right) + K_2 \left(1 + e_0 - \frac{1}{2} e_1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) \right]$$

$$t_{pr(st)} - \bar{Y} = \bar{Y} \left[(e_0) + K_1 \left(1 + e_0 - e_1 - e_0 e_1 + e_1^2 \right) + K_2 \left(1 + e_0 - \frac{1}{2} e_1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) \right]$$

Using the formula for MSE of the proposed and put the values

$$MSE(t_{pr(st)}) = E[t_{pr(st)} - \bar{Y}]^2$$

$$MSE(t_{pr(st)}) = \bar{Y}^2 E \left[e_0 + K_1 \left(1 + e_0 - e_1 - e_0 e_1 + e_1^2 \right) + K_2 \left(1 + e_0 - \frac{1}{2} e_1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) \right]^2$$

$$\begin{aligned} MSE(t_{pr(st)}) = \bar{Y}^2 \bigg\{ & E(e_0^2) + K_1^2 E(1 + 2e_0 - 2e_1 - 4e_0 e_1 + e_0^2 + 3e_1^2) \\ & + K_2^2 E(1 + 2e_0 - e_1 + e_0^2 - 2e_0 e_1 + e_1^2) \\ & + 2K_1 E(e_0 - e_0 e_1 + e_0^2) \\ & + 2K_1 K_2 E \left(1 + 2e_0 - \frac{3}{2} e_1 + e_0^2 - 3e_0 e_1 + \frac{15}{8} e_1^2 \right) \\ & + 2K_2 E \left(e_0 - \frac{1}{2} e_0 e_1 + e_0^2 \right) \bigg\} \end{aligned}$$

$$\begin{aligned} MSE(t_{pr(st)}) = \bar{Y}^2 E(e_0^2) + K_1^2 \bar{Y}^2 E(1 + 2e_0 - 2e_1 - 4e_0 e_1 + e_0^2 + 3e_1^2) \\ + K_2^2 \bar{Y}^2 E(1 + 2e_0 - e_1 + e_0^2 - 2e_0 e_1 + e_1^2) \\ + 2K_1 \bar{Y}^2 E(e_0 - e_0 e_1 + e_0^2) \\ + 2K_1 K_2 \bar{Y}^2 E \left(1 + 2e_0 - \frac{3}{2} e_1 - 3e_0 e_1 + e_0^2 + \frac{15}{8} e_1^2 \right) \\ + 2K_2 \bar{Y}^2 E \left(e_0 - \frac{1}{2} e_0 e_1 + e_0^2 \right) \end{aligned}$$

The MSE of proposed estimator is obtained as,

$$MSE(t_{pr(st)}) = \bar{Y}^2 \left[v_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R \right] \quad (16)$$

$$\begin{aligned} \text{where } A &= 1 + \omega_{2.0} - 4\omega_{1.1} + 3\omega_{0.2}, \quad B = 1 + \omega_{2.0} - 2\omega_{1.1} + \omega_{0.2}, \quad C = \omega_{2.0} - \omega_{1.1}, \\ Q &= 1 + \omega_{2.0} - 3\omega_{1.1} + \frac{15}{8}\omega_{0.2}, \quad R = \omega_{2.0} - \frac{1}{2}\omega_{1.1}, \quad \omega_{2.0} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2}, \\ \omega_{0.2} &= \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{ph}^2}{\bar{p}^2}, \quad \omega_{1.1} = \sum_{h=1}^L W_h^2 \lambda_h \frac{S_{yph}}{\bar{p}\bar{Y}}, \quad K_1 = \frac{QR-BC}{AB-Q^2}, \quad K_2 = \frac{CQ-AR}{AB-Q^2} \end{aligned}$$

Specific situations for the suggested estimator

In this section, we find same special cases of proposed estimator is as follow.

Specific situations: Case-I

Put $K_2 = 0$ in Eq. (15), we get

$$t'_{pr(st)} = \bar{y}_{st} + K_{11} \bar{y}_{st} \left[\frac{P}{p_{st}} \right] \quad (17)$$

Put the values of $\bar{y}_{st}$ and p_{st} in above equation and simplify, then we get

$$\begin{aligned} t'_{pr(st)} &= \bar{Y} \left[(1 + e_0) + K_{11} (1 + e_0 - e_1 - e_0 e_1 + e_1^2) \right] \\ t'_{pr(st)} - \bar{Y} &= \bar{Y} \left[e_0 + K_{11} (1 + e_0 - e_1 - e_0 e_1 + e_1^2) \right] \end{aligned} \quad (18)$$

Using the MSE formula and put the values

$$\begin{aligned} MSE(t'_{pr(st)}) &= E[t'_{pr(st)} - \bar{Y}]^2 \\ MSE(t'_{pr(st)}) &= \bar{Y}^2 E \left[e_0 + K_{11} (1 + e_0 - e_1 - e_0 e_1 + e_1^2) \right]^2 \\ MSE(t'_{pr(st)}) &= \bar{Y}^2 \left\{ E(e_0^2) + K_{11}^2 E(1 + 2e_0 - 2e_1 - 4e_0 e_1 + e_0^2 + 3e_1^2) + 2K_{11} E(e_0 - e_0 e_1 + e_0^2) \right\} \end{aligned}$$

Then the mean square error of first case is obtained as

$$MSE(t'_{pr(st)}) = \bar{Y}^2 [\omega_{2.0} + K_{11}^2 A + 2K_{11} C] \quad (19)$$

where, $K_{11} = -\left[\frac{C}{A}\right]$ is the optimum value.

Specific situations: Case-II

Put $K_1 = 0$ in Eq. (15), we get

$$t''_{pr(st)} = \bar{y}_{st} + K_{12} \bar{y}_{st} \exp \left[\frac{P - p_{st}}{P + p_{st}} \right] \quad (20)$$

Put the values of $\bar{y}_{st}$ and p_{st} in above equation, we get

$$\begin{aligned} t''_{pr(st)} &= \bar{Y} \left[(1 + e_0) + K_{12} \left(1 + e_0 - \frac{1}{2} e_1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) \right] \\ t''_{pr(st)} - \bar{Y} &= \bar{Y} \left[e_0 + K_{12} \left(1 + e_0 - \frac{1}{2} e_1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) \right] \end{aligned} \quad (21)$$

When we square and consider expectations on both sides of above equation, we obtain

$$\begin{aligned} MSE(t''_{pr(st)}) &= E[t''_{pr(st)} - \bar{Y}]^2 \\ MSE(t''_{pr(st)}) &= \bar{Y}^2 E \left[e_0 + K_{12} \left(1 + e_0 - \frac{1}{2} e_1 - \frac{1}{2} e_0 e_1 + \frac{3}{8} e_1^2 \right) \right]^2 \\ MSE(t''_{pr(st)}) &= \bar{Y}^2 \left\{ E(e_0^2) + K_{12}^2 E \left(1 + 2e_0 - e_1 + e_0^2 - 2e_0 e_1 + e_1^2 \right) + 2K_{12} E \left(e_0 - \frac{1}{2} e_0 e_1 + e_0^2 \right) \right\} \\ MSE(t''_{pr(st)}) &= \bar{Y}^2 E(e_0^2) + K_{12}^2 \bar{Y}^2 E \left(1 + 2e_0 - 2e_1 - e_0 e_1 + 3e_1^2 \right) + 2K_{12} \bar{Y}^2 E \left(e_0 - \frac{1}{2} e_0 e_1 + e_0^2 \right) \end{aligned}$$

Then the MSE is

$$MSE(t''_{pr(st)}) = \bar{Y}^2 [\omega_{2.0} + K_2^2 B + 2K_2 R] \quad (22)$$

where $K_{12} = -\left[\frac{R}{B}\right]$ is the optimum value.

Efficiency comparisons

The robustness of the proposed estimator has been examined by analyzing its sensitivity to variations in the auxiliary attribute and potential model misspecification, showing that it maintains stable performance under moderate deviations from underlying assumptions. Its asymptotic behavior has also been studied, establishing consistency and efficiency as the sample size increases. Furthermore, the optimality conditions have been analytically interpreted to clearly identify the parameter settings and sampling scenarios under which the estimator achieves minimum MSE and superior efficiency compared to existing methods.

The proposed estimator $t_{pr(st)}$ is said to be more efficient than an existing estimator $t_{f(st)}$ for all $f = 1, 2, \dots, \text{infinite}$, if its MSE is smaller, that is,

$$MSE(t_{pr(st)}) < MSE(t_{f(st)})$$

Specific Efficiency Conditions: Accordingly, the proposed estimator is more efficient than:

1. The stratified sample mean

$$MSE(t_{pr(st)}) < MSE(t_{1(st)})$$

that is

$$v_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R < v_{2.0}$$

2. Estimator $t_{2(st)}$

$$MSE(t_{pr(st)}) < MSE(t_{2(st)})$$

that is

$$v_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R < \omega_{2.0} - 2\omega_{1.1} + \omega_{0.2}$$

3. Estimator $t_{3(st)}$

$$MSE(t_{pr(st)}) < MSE(t_{3(st)})$$

that is

$$\nu_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R < \omega_{2.0} - \omega_{1.1} + \frac{1}{4} \omega_{0.2}$$

4. Estimator $t_{4(st)}$

$$MSE(t_{pr(st)}) < MSE(t_{4(st)})$$

that is

$$\nu_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R < \omega_{2.0} - 2\delta_1 \omega_{1.1} + \delta_1^2 \omega_{0.2}$$

5. Estimator $t_{5,i(st)}$

$$MSE(t_{pr(st)}) < MSE(t_{5,i(st)}), \text{ for all } i = 1, 2, \dots, 10.$$

that is

$$\begin{aligned} \nu_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R &< k^2 \omega_{2.0} \\ &+ \varrho_i^2 (3k^2 - 2k) \omega_{0.2} - 2\varrho_i (3k^2 - 2k) \omega_{1.1} + (k - 2)^2 \\ \nu_{2.0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R &< k^2 \omega_{2.0} \\ &+ \varrho_i^2 (3k^2 - 2k) \omega_{0.2} - 2\varrho_i (3k^2 - 2k) \omega_{1.1} + (k - 2)^2 \end{aligned}$$

6. Estimator $t_{6(st)}$

$$MSE(t_{pr(st)}) < MSE(t_{6(st)})$$

that is

Table 1 Summary of the real data

h	N_h	n_h	$\bar{Y}_h$	P_h	S_{yh}	S_{ph}	S_{yph}	ρ_h
1	106	13	1536.09	0.25	6425.09	0.43	996.59	0.3
2	106	24	2212.54	0.29	11551.54	0.46	1404.01	0.2
3	94	55	9384.31	0.47	29907.48	0.51	48674.99	0.3
4	171	95	5588.01	0.49	28643.78	0.50	-2743.02	0.1
5	204	10	966.96	0.36	2389.77	0.48	449.02	0.3
6	173	3	404.39	0.12	945.75	0.32	204.18	0.5

Table 2 Summary of the real data

h	N_h	n_h	$\bar{Y}_h$	P_h	S_{yh}	S_{ph}	S_{yph}	ρ_h
1	127	31	707.74	0.952	883.84	0.21	25.27	0.13
2	117	21	413.01	0.974	944.92	0.15	9.98	0.06
3	103	29	573.17	0.932	1033.47	0.25	37.45	0.14
4	170	38	424.66	0.888	810.58	0.32	44.62	0.17
5	205	22	267.03	0.912	403.65	0.28	21.04	0.18
6	201	39	393.84	0.950	711.72	0.22	18.66	0.12

Table 3 MSE and PRE for the proposed and existing estimators

Estimators	MSE		PRE	
	Population I	Population II	Population I	Population II
$t_{1(st)}$	387840.031	2528.567	100.000	100
$t_{2(st)}$	488061.170	2587.625	79.465	97.71768
$t_{3(st)}$	370610.637	2414.265	104.649	104.7344
$t_{4(st)}$	370206.233	2468.956	104.763	102.4144
$t_{5,1(st)}$	946008.974	2489.654	40.998	102.0278
$t_{5,2(st)}$	1167128.568	2478.312	33.230	101.3094
$t_{5,3(st)}$	1024536.500	2495.887	37.855	102.3902
$t_{5,4(st)}$	959449.333	2469.541	40.423	101.7982
$t_{5,5(st)}$	1539604.715	2483.902	25.191	102.2801
$t_{5,6(st)}$	1197106.985	2472.198	32.398	98.95975
$t_{5,7(st)}$	955637.321	2555.147	40.584	102.9601
$t_{5,8(st)}$	1067888.297	2455.871	36.318	102.6033
$t_{5,9(st)}$	967272.468	2464.412	40.096	104.4921
$t_{5,10(st)}$	1096245.548	2419.864	35.379	100.5344
$t_{6(st)}$	899947.154	2515.125	43.096	119.0159
$t_{7(st)}$	352750.894	2124.562	109.947	120.1725
$t'_{pr(st)}$	346909.714	2104.114	111.799	117.8477
$t''_{pr(st)}$	386490.790	2145.622	111.934	125.0983
$t_{pr(st)}$	314577.953	2021.264	123.289	102.0278

$$\nu_{2,0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R < \omega_{2,0} + \zeta_1^2 (4\alpha^2 - 4\alpha + 1) \omega_{0,2} - 2\zeta_1 (2\alpha - 1) \omega_{1,1}$$

7. Estimator $t_{7(st)}$

$$MSE(t_{pr(st)}) < MSE(t_{7(st)})$$

that is

$$\nu_{2,0} + K_1^2 A + K_2^2 B + 2K_1 C + 2K_1 K_2 Q + 2K_2 R < a + z_1^2 b + z_2^2 c + 2z_1 d + 2z_2 p + 2z_1 z_2 q$$

These conditions show that the proposed estimator outperforms existing estimators whenever the chosen constants satisfy the above inequalities. By appropriately

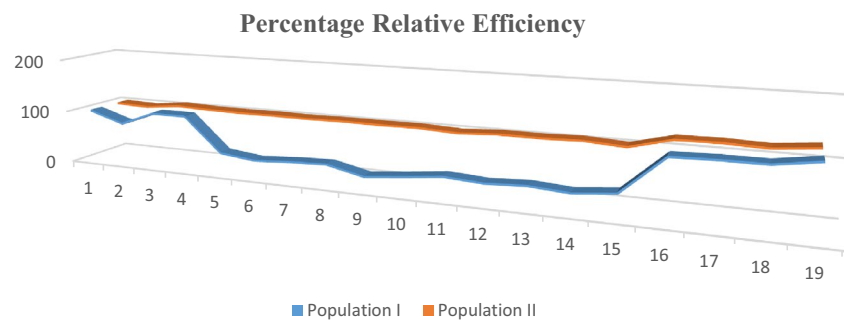


Fig. 1 PREs for the proposed and existing estimators

Table 4 Summary of the simulation data

h	N_h	n_h	$\bar{Y}_h$	P_h	S_{yh}^2	S_{ph}^2	S_{yph}
1	2000	1500	30.27	30.59	431.12	1075.64	-14.11
2	2500	1200	37.89	38.62	582.47	1698.16	-0.44
3	3000	1800	45.05	45.58	681.06	2462.87	26.61
4	2500	1500	52.31	52.01	761.23	3644.83	-10.56

selecting these constants, the MSE of the proposed estimator can be minimized, ensuring superior efficiency across a wide range of practical situations.

Empirical study: using two real population

Population I: In this study, the data set is taken from Akintunde et al. [10]. The quantity of apples produced in 1999 is considered as the study variable, while the number of apple trees in 1999 is used as the auxiliary variable. The data were obtained from the Turkish Republic Institute of Statistics. For stratification purposes, Turkey was divided into six regions: (1) Marmara, (2) Aegean, (3) Mediterranean, (4) Central Anatolia, (5) Black Sea, and (6) East and Southeast Anatolia. Random samples were selected from each stratum. The sample sizes for each stratum were determined using the Neyman allocation method as described in the original data source. The real data set summary is given in Table 1.

Now we find the numerical value of the MSE and PRE of the proposed estimator and the same existing estimator for data sets taken from Table 1.

Population II: To illustrate the efficiency of the suggested estimators in the application, we consider the data set reported in Cochran [27], where the number of teachers is taken as the study variable (y) and the number of students is used as the auxiliary attribute (p). The Neyman allocation method is used to allocate the sample sizes to different strata. The summary statistics of the data are presented in Table 2.

Table 3 presents the MSE and PRE of the proposed estimator in comparison with existing estimators under the given stratified sampling framework. The MSE is used as a measure of precision, while the PRE facilitates a relative comparison of efficiency by taking a suitable estimator as the reference. A lower MSE and higher PRE indicate better performance of an estimator. The results reported in this table demonstrate the

effectiveness and superiority of the proposed estimator over the competing estimators for the considered dataset.

Table 3 show the percentage relative efficiency (PRE) of the proposed estimator in comparison with several existing estimators under stratified sampling. It is evident from the table that the proposed estimator consistently attains the highest PRE among all estimators considered, indicating a substantial improvement in efficiency. These findings confirm the theoretical results and demonstrate the practical superiority of the proposed estimator for population mean estimation in stratified sampling.

Figure 1, the percentage relative efficiency (PRE) of the proposed estimator is observed to be the highest among all estimators considered for the population. This demonstrates the superior performance of the proposed estimator in comparison with the existing estimators. The estimators $t_{1(st)}$, $t_{2(st)}$, $t_{3(st)}$, $t_{4(st)}$, $t_{5,i(st)}$, $t_{6(st)}$, $t'_{pr(st)}$, $t''_{pr(st)}$ and $t_{pr(st)}$ are displayed from left to right in the figure.

Empirical study: simulation data

To examine the performance and robustness of the proposed estimators under realistic conditions, a sensitivity analysis was conducted using a non-normal stratified population. The total population size was fixed at 10,000 units and divided into four unequal strata. A stratified random sample of size 6,000 was selected using non-proportional allocation, and simple random sampling without replacement was applied within each stratum. The study variable y was generated from a Gamma distribution, while the auxiliary variable x followed a Lognormal distribution to incorporate skewness in the generated data. Specifically, the variables were generated as

$y \sim \text{Gamma}(\alpha = 4, \beta = 8)$, $x \sim \text{Lognormal}(\mu = 3.5, \sigma^2 = 0.6)$ where α and β represent the shape and scale parameters of the Gamma distribution, and μ and σ^2 denote

Table 5 MSE and PRE for the proposed and existing estimators of the simulation data

Estimators	MSE	PRE
$t_{1(st)}$	0.04495	100.0000
$t_{2(st)}$	0.20543	21.8809
$t_{3(st)}$	0.24587	18.2820
$t_{4(st)}$	0.15487	29.0243
$t_{5,1(st)}$	0.04587	97.9943
$t_{5,2(st)}$	0.05974	75.2427
$t_{5,3(st)}$	0.07512	59.8376
$t_{5,4(st)}$	0.01452	309.5730
$t_{5,5(st)}$	0.06847	65.6492
$t_{5,6(st)}$	0.02651	169.5587
$t_{5,7(st)}$	0.03452	130.2144
$t_{5,8(st)}$	0.08975	50.0836
$t_{5,9(st)}$	0.04571	98.3373
$t_{5,10(st)}$	0.09854	45.6160
$t_{6(st)}$	0.05245	85.7007
$t_{7(st)}$	0.02784	161.458
$t'_{pr(st)}$	0.01248	360.1763
$t''_{pr(st)}$	0.02451	183.3945
$t_{pr(st)}$	0.007845	572.9764

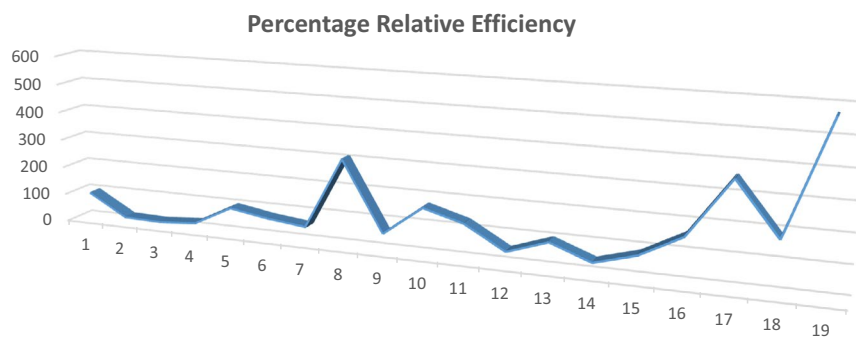


Fig. 2 PRE for the proposed and existing estimators

the mean and variance parameters of the Lognormal distribution. Stratum-wise population and sample statistics were then computed to assess efficiency and robustness given in Table 4.

A sensitivity analysis was carried out to evaluate the robustness and stability of the proposed estimators under varying distributional and sampling conditions. Specifically, key parameters such as the degree of skewness in the study variable y (modeled via different shape and scale parameters of the Gamma distribution) and the auxiliary variable x (through varying parameters of the Lognormal distribution) were systematically altered. In addition, different non-proportional allocation schemes and stratum sample sizes were considered to assess the impact of sampling design on estimator performance. The results indicate that the proposed estimators maintain lower mean squared error and higher efficiency across a wide range of scenarios, demonstrating their robustness and reliability even under departures from normality and changes in population structure.

The results presented in Table 5 show the percentage relative efficiency (PRE) of the proposed estimator in comparison with several existing estimators under stratified sampling. It is evident from the table that the proposed estimator consistently attains the highest PRE among all estimators considered, indicating a substantial improvement in efficiency. These findings confirm the theoretical results and demonstrate the practical superiority of the proposed estimator for population mean estimation in stratified sampling.

Figure 2, the percentage relative efficiency (PRE) of the proposed estimator is observed to be the highest among all estimators considered for the population. This demonstrates the superior performance of the proposed estimator in comparison with the existing estimators. The estimators $t_{1(st)}$, $t_{2(st)}$, $t_{3(st)}$, $t_{4(st)}$, $t_{5,i(st)}$, $t_{6(st)}$, $t'_{pr(st)}$, $t''_{pr(st)}$ and $t_{pr(st)}$ are displayed from left to right in the figure.

Results and discussion

The empirical and simulation results presented in Tables 3 and 5 provide a comprehensive comparison of the proposed estimators under stratified sampling using auxiliary attribute information. The performance of the estimators is evaluated in terms of mean squared error (MSE) and percentage relative efficiency (PRE), taking the usual

stratified estimator as the benchmark. From Table 3, the proposed estimator achieves the lowest mean squared error and the highest percentage relative efficiency, demonstrating clear superiority over the usual stratified estimator and other competitors. The modified proposed estimators show only marginal efficiency gains, with percentage relative efficiency values slightly above 100%. Among the traditional estimators, the third and fourth estimators perform moderately better than the usual estimator, while the fifth and sixth estimators exhibit substantially higher mean squared errors and low efficiency, with percentage relative efficiency below 45%. These results indicate that efficiency gains from auxiliary attributes depend critically on the appropriate choice of estimator form. The simulation results in Table 5 clearly support the empirical findings. The proposed estimator performs best, with the smallest mean squared error and the highest percentage relative efficiency, showing a substantial improvement in efficiency. Although the modified proposed estimators and a few existing methods show some improvement over the usual estimator, their performance remains well below that of the proposed approach, while the remaining estimators perform poorly under the simulated setting. Overall, both empirical and simulation results confirm that the proposed estimators are highly efficient for estimating the population mean under stratified sampling with auxiliary attribute information. The consistency of the results demonstrates their robustness and practical applicability. The improved performance is due to the effective utilization of auxiliary attributes and the optimized structure of the estimators. However, the efficiency of these estimators depends on the strength of the relationship between the study variable and auxiliary attributes, and the use of first-order approximations may be less accurate in cases of small sample sizes or highly skewed populations.

Conclusion

In this study, an improved estimator for the population mean in stratified sampling was proposed by making effective use of auxiliary attribute information. The motivation behind this work was to enhance estimation accuracy, especially in practical situations where such auxiliary information is readily available but often underutilized. The proposed estimator was developed with a focus on its application in agricultural studies, particularly in the context of apple production. First-order approximations were used because they provide simple and tractable expressions for bias and mean squared error, making the analysis and comparison of estimators easier. They are sufficiently accurate for moderate to large sample sizes and are widely accepted as a standard approach in sampling theory, whereas higher-order terms introduce substantial complexity with limited practical benefit. Expressions for bias and mean squared error (MSE) were derived to understand the statistical properties of the estimator, and the conditions under which it performs better than existing methods were identified. The theoretical development was supported by both real data analysis and simulation studies. The results consistently showed that the proposed estimator provides more precise and efficient estimates compared to traditional estimators used in stratified sampling.

An important observation from this study is the strong agreement between the theoretical results and empirical findings, highlighting the reliability and stability of the proposed approach. This demonstrates that the estimator is not only mathematically sound but

also practically applicable. Overall, the study shows that incorporating auxiliary attribute information can substantially improve the estimation of population means, particularly in agricultural surveys such as apple production. Future research may extend the proposed estimator to more complex sampling designs, multiple auxiliary attributes, and applications in environmental, health, and socio-economic surveys.

Acknowledgements

The author expresses their heartfelt gratitude to the learned referee and the editor of the journal for the constructive comments that improved the earlier draft.

Author contributions

All authors contributed significantly to the development of this manuscript. M.K.V. and S.A.K. were involved in conceptualization, methodology, formal analysis, and writing of the original draft. Both authors contributed to software, validation, investigation, data curation, supervision, resources, and writing—review and editing. All authors reviewed and approved the final version of the manuscript.

Funding

Open access funding provided by Manipal Academy of Higher Education, Manipal.

Data availability

The data used in this study were obtained from Akintunde et al. [10] and also simulated data generated for methodological validation.

Declarations

Ethics approval and consent to participate

The authors declare that the study did not involve experiments on humans or animals.

Competing interests

The authors declare no competing interests.

Received: 22 May 2026 Accepted: 9 August 2026

Published online: 21 September 2026

References

1. Kaur P. On the estimation of population mean in stratified sampling. *Biom J.* 1985;27(1):101–5.
2. Bahl S, Tuteja RK. Ratio and product type estimator. *J Inf Optim Sci.* 1991;9:159–63.
3. Diana G. A class of estimators of the population mean in stratified random sampling. *Statistica (Bologna).* 1993;53(1):59–66.
4. Kadilar C, Cingi H. Ratio estimators in stratified random sampling. *Biom J.* 2003;45(2):218–25.
5. Koyuncu N, Kadilar C. Ratio and product estimators in stratified random sampling. *J Stat Plan Inference.* 2009;139(8):2552–8.
6. Malik S, Singh R. Some improved multivariate-ratio-type estimators using geometric and harmonic means in stratified random sampling. *Int Sch Res Not.* 2012;2012(1):509186.
7. Lu J. Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural, biomedical, and power engineering. *Math Probl Eng.* 2017;2017(1):8704734.
8. Zaman T. Efficient estimators of population mean using auxiliary attribute in stratified random sampling. *Adv Appl Stat.* 2019;56(2):153–71.
9. Zaman T, Kadilar C. On estimating the population mean using auxiliary character in stratified random sampling. *J Stat Manag Syst.* 2020;23(8):1415–26.
10. Akintunde AA, Olayiwola OM, Adewole AI, Ojenike OT, Igbalajobi MM. An improved estimation of population mean using auxiliary attribute in stratified sampling. *Niger J Pure Appl Sci.* 2023;36(1):4463–71.
11. Zaagan AA, Meetei MZ, Singh S, Gupta R, Yadav SK, Verma MK. Computing the population mean in stratified sampling using an auxiliary attribute. *J Auton Intell.* 2024;7(5):12–21.
12. Daraz U, Almulhim FA, Alomair MA, Alomair AM. Population median estimation using auxiliary variables: a simulation study with real data across sample sizes and parameters. *Mathematics.* 2025;13(10):1660.
13. Daraz U, Wu J, Agustiana D, Emam W. Finite population variance estimation using Monte Carlo simulation and real life application. *Symmetry.* 2025;17(1):84.
14. Daraz U, Aljohani HM, Alshanbari HM. Efficient median estimation for stratified multi-population data: health services, medical workforce, and medical education. *BMC Med Res Methodol.* 2026. <https://doi.org/10.1186/s12874-026-02829-w>.
15. Verma MK, Vishwakarma GK, Yadav SK. Uncertainty-based strategy for population mean estimation using stratified sampling. *Commun Stat Case Stud Data Anal Appl.* 2025. <https://doi.org/10.1080/23737484.2025.2591905>.
16. Verma MK, Yadav SK, Varshney R, Vishwakarma GK. Improved estimation strategy of mean using linear cost function under stratified sampling. *Qual Quant.* 2025;59(Suppl 2):1143–62.
17. Verma MK. Optimal estimation of two population means under stratified sampling. *Qual Quant.* 2025. <https://doi.org/10.1007/s1135-025-02469-0>.

18. Yadav SK, Kumar Verma M, Varshney R. Optimal strategy for elevated estimation of population mean in stratified random sampling under linear cost function. *Ann Data Sci.* 2025;12(2):517–38.
19. Chaudhary MK, Abdelwahab MM, Bhardwaj N, Hasaballah MM. Estimation through calibration under stratified sampling with non-response and measurement error effects. *Math.* 2026;14(3):439.
20. Kumar A, Bhushan S, Pokhrel R. Direct and synthetic logarithmic estimators in small area estimation via ranked set sampling. *Stat.* 2026;60(1):70–91.
21. Robbenolt D, Pendyala RM, Bhat CR. Data collection, weighting, and modeling techniques to estimate consistent population parameters. *Transp Res B Methodol.* 2026;203:103349.
22. Verma MK, Yadav SK, Varshney R, Vishwakarma GK. An optimal strategy for estimating the population mean in stratified random sampling under an auxiliary attribute. *Life Cycle Reliab Saf Eng.* 2026;15(1):69–75.
23. Verma MK, Yadav SK, Yadav R, Yadav SK. Neutrosophic approaches to population mean estimation in stratified sampling: application to the Indian temperature records. *Qual Quant.* 2026. <https://doi.org/10.1007/s11135-026-02621-4>.
24. Verma MK, Sharma SR, Varshney R, Yadav SK. On optimal estimation in supply chain management using stratified sampling. *Qual Quant.* 2026. <https://doi.org/10.1007/s11135-026-02733-x>.
25. Verma MK, Varshney R, Yadav SK, Kumari P. An improved estimation technique for the population mean in post-stratified sampling: applications and simulation-based evaluation. *Qual Quant.* 2026. <https://doi.org/10.1007/s11135-026-02675-4>.
26. Naik VD, Gupta PC. A note on estimation of mean with known population proportion of an auxiliary character. *J Indian Soc Agric Stat.* 1996;48(2):151–8.
27. Cochran WG. *Sampling techniques*. 3rd ed. Hoboken: Wiley eastern limited; 1977. p. 325.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.