Journal of Autonomous Intelligence (2024) Volume 7 Issue 5.
doi:10.32629/jai.v7i5.1672

Original Article

Computing the Population Mean in Stratified Sampling Using an Auxiliary Attribute

Abdullah A Zaagan[1], Mutum Zico Meetei[1], Shreyanshu Singh[2], Rajesh Gupta[3*], Subhash Kumar Yadav[4], Mukesh Kumar Verma[4*]

[1]Department of Mathematics, College of Science, Jazan University, P.O. Box 114, Jazan 45142, Saudi Arabia.

Email: azaagan@jazanu.edu.sa; mmeetei@jazanu.edu.sa

Orcid ID: https://orcid.org/0009-0009-6838-6521; https://orcid.org/0000-0002-9168-5126

[2]School of Management, Babu Banarasi Das University, Lucknow - 226028, UP, India, shreyanshu_7@bbdu.ac.in, Orcid ID: 0000-0001-8745-6227

[3]Global Data insight and Analysis (GDIA), Ford Motor Company, American Rd, Dearborn, MI 48126, USA, rajeshgupta2004@gmail.com

[4]Department of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow-226025, India

Email: drskystats@gmail.com; vermamukeshkumar968@gmail.com

ABSTRACT

This paper proposes a novel ratio type estimator for the population mean using an auxiliary attribute by implying one auxiliary variable in stratified random sampling using conventional product, exponential, and logarithmic ratio type estimators. The proposed estimator's MSE and PRE are determined, and PRE is compared with existing estimators. The proposed estimator is more effective than other existing estimators according to theoretical observations, which verifies its numerical examples. The proposed estimator may be used for practical applicability in real life, including agricultural sciences, biological sciences, commercial sciences, economic sciences, engineering sciences, medical sciences, social sciences etc.

Keywords: *Study Variable; Auxiliary Attribute; Ratio Type Estimators; PRE*

ARTICLE INFO
Received: Mar 27, 2024
Accepted: May 7, 2024
Available online: May 24, 2024

***CORRESPONDING AUTHOR**
Rajesh Gupta[1], Mukesh Kumar Verma[2].
[1]Global Data insight and Analysis (GDIA), Ford Motor Company, American Rd, Dearborn, MI 48126, USA.
[2]Department of Statistics Babasaheb Bhimrao Ambedkar University, Lucknow-226025, India.

CITATION
Abdullah A Zaagan, Mutum Zico Meetei, Shreyanshu Singh, Rajesh Gupta, Subhash Kumar Yadav, Mukesh Kumar Verma. Computing the population mean in stratified sampling using an auxiliary attribute. Journal of Autonomous Intelligence 2024; 7(5): 12-21. doi: 10.32629/jai.v7i5.1672

1. Introduction

In large populations, sampling is a simple, accurate, and reliable procedure that provides valuable information that allowing for decision-making based on the characteristics of the population. The sample provides the most comprehensive representation of a large population. The main goal of this study is to search for the estimator, whose sampling distribution is close to real population mean, incorporating auxiliary data that can either positively or negatively impact the main variable. Various fields of study, including population studies, marketing, agriculture, economics, industry, medicine, and social science, use these estimators. Simple random sampling, systematic sampling, stratified sampling, cluster sampling, double sampling, two-stage sampling, probability proportional to size sampling, and multi-stage sampling are only a few of the numerous sampling techniques available.

Dividing the whole heterogeneous population into different strata before sampling enhances the precision of conclusions drawn from the collected data. Auxiliary variables, primarily ratio, product, and exponential estimators, have been employed to increase the accuracy of estimators. Using stratified sampling makes it possible to partition the entire heterogeneous population into several strata to maintain the homogeneity within each stratum along with feasibility. Sampling from each stratum is usually done using a simple random sampling technique. Stratification is a technique used in modern surveys to improve the precision of estimations. Cochran[1] introduced the ratio technique for estimation using auxiliary data and a traditional ratio estimator for positive correlation between auxiliary and primary variables. Srivastava[2] developed a class of population mean estimators using the auxiliary variable's known population mean. Sisodia *et al.*[3] used the supplementary data to estimate the population mean. A variety of well-known estimators, including ratio, product, linear regression, and ratio type, are accessible for improved estimates.

A population could be divided into two classes proportional to an attribute, equivalent to simple random sampling. Therefore, based on whether they have the particular attribute or not, the population's units are divided into these two varieties. The population proportion of the defined attribute may be of interest for our purposes after we obtain a sample of size .

The exponential estimator was introduced by Bahl and Tuteja[4] in Simple Random Sampling. Using the two-phase sampling, Singh and Vishwakarma[5] utilized known values of a particular population parameters. A family of exponential estimators have been demonstrated by Koyuncu[6] to estimate the population mean using the information of two auxiliary attributes. Haq and Shabbir[7] proposed two improved estimator families for predicting the finite population mean. Sanaullah *et al.*[8] proposed exponential estimators in the stratified two-phase sampling. Exponential estimators can also be used to the exponential relationship between the auxiliary attribute and study variable. For simple random and stratified random samplings, Shabbir and Gupta[9] proposed an estimator that had been more effective than various competing estimators. A novel approach to measuring population variance in simple random sampling has been given by Lone and Tailor[10]. In their general form, Cekim and Kadilar[11] improved two classes of estimators for estimating the population mean in stratified random sampling. Yadav *et al.*[12] enhanced the estimation of the population mean by utilizing several qualities of auxiliary information. Zaman[13] introduced the efficient estimators of population mean using auxiliary attributes in stratified random sampling. Tolga and Kadilar[14] proposed the general category of exponential estimators. Yadav *et al.*[15] suggested a memory-type product estimator to estimate the population mean of the study's variable in stratified sampling and Yadav *et al.*[16] also proposed generalized ratio-cum-exponential-log ratio type estimators of population mean under a simple random sampling scheme. Akintunde *et al.*[17] examine an improved estimation of population mean using auxiliary attribute in stratified sampling. Yadav *et al.*[18] introduced optimal strategy for elevated estimation of population mean in stratified random sampling under linear cost function. Koyuncu and Kadilar[19] proposed efficient estimators for the population mean.

Mathew *et al.*[20] introduced the procedure of neyman allocation in stratified random sampling. Singh and Solanki[21] introduced improved estimation of population mean in simple random sampling using information on auxiliary attribute. Singh *et al.*[22] introduced the effective class of estimators for the finite population mean in stratified random sampling using an auxiliary attribute. Koyuncu[23] proposed the families of estimators for population mean using information on auxiliary attribute in stratified random sampling. Ahmadini *et al.*[24] proposed the restructured searls family of estimators of population mean in the presence of nonresponse.

This paper suggests a type of estimator named product-cum-logarithmic ratio for estimating the population mean in stratified random sampling, utilizing information on an auxiliary attribute. The proposed estimator's bias and Mean Squared Error (MSE) are derived for an approximation of order one and it has been compared with competing estimators. The effectiveness of the proposed estimator over the existing ones is theoretically validated, and a numerical example illustrates its performance.

2. Notations

Consider a finite population of size N and random sample of size n_h is drawn using simple random sampling without replacement (SRSWOR) from the h^{th} stratum such that it is divided into L strata. Let $\emptyset$ be an auxiliary attribute and Y be the study variable taking values $\emptyset_{hi}$ and y_{hi} respectively $(h = 1, 2, ..., L;\ i = 1, 2, ..., N_h)$ on i^{th} unit of the h^{th} stratum. Each stratum provides a sample of size n_h, to get a sample of size n. With $i = 1,2, ... N_h$ and $h = 1,2, ... L$, let y_{hi} and $\emptyset_{hi}$ represent the observed values of study variable Y and auxiliary attribute $\emptyset$ respectively, on the i^{th} unit of the h^{th} stratum. Assume further that either the presence or absence of an attribute $\emptyset_h$ implies a complete dichotomy in the population considered and there are only two possibilities, 1 and 0 (1 if i^{th} unit of population possesses attribute $\emptyset$ and 0 if it does not possess it) values for the attribute $\emptyset_h$. Let , $\mathrm{A} = \sum_{i=1}^{N} \emptyset_i$, $\mathrm{A}_h = \sum_{i=1}^{N} \emptyset_{hi}$, $a = \sum_{i=1}^{n_h} \emptyset_i$ and $a_h = \sum_{i=1}^{n_h} \emptyset_{hi}$. Represent the total number of units with attribute ϕ in the population, sample, population stratum h, and sample, respectively. Let $\mathrm{P} = \frac{\mathrm{A}}{N}$, $\mathrm{P}_h = \frac{A_h}{N_a}$, $p = \frac{a}{n}$and $a_h = \frac{a_h}{n_h}$. The followings may be considered as, $\bar{Y}_h$ is h^{th} stratum mean for the studied variable and $\emptyset_h$ is h^{th} stratum 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 $p = \mathrm{P}(1 + e_1)$, where e_0 and e_1 are errors of approximation and the value expectation of e_0 and e_1, we have

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

$$E(e_0^2) = \sum_{h=1}^{L} W_h^2 \lambda_h \frac{S_{yh}^2}{\bar{Y}^2} = \vartheta_{2.0}$$

$$E(e_1^2) = \sum_{h=1}^{L} W_h^2 \lambda_h \frac{S_{ph}^2}{\mathrm{P}^2} = \vartheta_{0.2}$$

$$E(e_0 e_1) = \sum_{h=1}^{L} W_h^2 \lambda_h \frac{S_{yph}}{\mathrm{P}\bar{Y}} = \vartheta_{1.1} \qquad (1)$$

Where,

$$\lambda_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) ; h = 1,2, ..., L$$

and

$$\upsilon_{a.b} = \sum_{h=1}^{L} W_h^{a+b} \lambda_h \sum_{i=1}^{N_h} \frac{(y_{hi} - \bar{Y}_h)^a (p - \mathrm{P})^b}{\bar{Y}^a \mathrm{P}^b} \qquad (2)$$

S_{yh}^2 and S_{ph}^2 represents the study variable's and auxiliary attribute in population variances, and S_{yph}^2 represents the covariance of study variable and auxiliary attribute for the population in the h^{th} stratum.

3. Estimators in Sampling Literature

Using a single auxiliary characteristic, the traditional combined ratio estimator in stratified random sampling is as follows

$$\boldsymbol{t_{1(st)}} = \bar{\boldsymbol{y}}_{\boldsymbol{st}} \left[\frac{\mathrm{P}}{P_{st}}\right] \qquad (3)$$

The combined ratio estimator's variance

$$MSE(t_{1(st)}) = \bar{Y}^2[\upsilon_{2.0} - 2\upsilon_{1.1} + \upsilon_{0.2}] \qquad (4)$$

The following exponential estimator has been suggested by Sharma and Singh (2013).

$$\boldsymbol{t_{2(st)}} = \bar{\boldsymbol{y}}_{\boldsymbol{st}} exp\left[\frac{\mathrm{P} - \mathrm{p}_{st}}{\mathrm{P} + \mathrm{p}_{st}}\right] \qquad (5)$$

The bias and MSE, respectively

$$Bias(t_{3(st)}) = \bar{Y}\left[\frac{3}{8}\upsilon_{0.2} - \frac{1}{2}\upsilon_{1.1}\right]$$

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

Zaman[13] has been introduced the efficient estimators of population mean using auxiliary attribute in stratified random sampling is,

$$\boldsymbol{t_{3(st)}} = \bar{\boldsymbol{y}}_{\boldsymbol{st}} exp\left[\frac{(\mathrm{kp}_{st} + l) - (\mathrm{kp}_{st} + l)}{(\mathrm{kp}_{st} + l) + (\mathrm{kp}_{st} + l)}\right] \qquad (7)$$

The bias and MSE, respectively

$$Bias(t_{3(st)}) = \bar{Y}[\upsilon_{0.2} - \upsilon_{1.1}]$$

$$MSE(t_{3(st)}) = \bar{Y}^2[\upsilon_{2.0} - 2\delta_i \upsilon_{1.1} + \delta_i^2 \upsilon_{0.2}] \qquad (8)$$

Where

$$\delta_1 = \frac{1}{2} \quad \delta_2 = \frac{\mathrm{P}}{2(\mathrm{P} + \beta_2)},\ \delta_3 = \frac{\mathrm{P}}{2(\mathrm{P} + C_p)},\ \delta_4 = \frac{\mathrm{P}}{2(\mathrm{P} + \rho_p)},\ \delta_5 = \frac{\beta_2 \mathrm{P}}{2(\beta_2 \mathrm{P} + C_p)},\ \delta_6 = \frac{C_p \mathrm{P}}{2(\beta_2 + C_p \mathrm{P})},$$

$$\delta_7 = \frac{\beta_2 \mathrm{P}}{2(\rho_p + C_p \mathrm{P})},\ \delta_8 = \frac{\rho_p \mathrm{P}}{2(\rho_p \mathrm{P} + C_p)},\ \delta_9 = \frac{\beta_2 \mathrm{P}}{2(\beta_2 \mathrm{P} + \rho_p)},\ \delta_{10} = \frac{\rho_p \mathrm{P}}{2(\rho_p \mathrm{P} + \beta_2)},$$

Inspired by Koyuncu and Kadilar[19], Zaman and Kadilar[14] improved the general category of exponential estimators in the following ways:

$$t_{4,i(st)} = k\bar{y}_{st}exp\left[\frac{(kp_{st}+l)-(kp_{st}+l)}{(kp_{st}+l)+(kp_{st}+l)}\right] \quad (9)$$

where k is an optimal constant that will be discovered later. In this case,

The bias and MSE, respectively

$$Bias(t_{4i(st)}) = \bar{Y}[\varrho_i^2 \upsilon_{0.2} - \delta_i \upsilon_{1.1}]$$

$$MSE(t_{4i(st)}) = \bar{Y}^2[k^2\upsilon_{2.0} + \varrho_i^2(3k^2-2k)\upsilon_{0.2} - 2\varrho_i(3k^2-2k)\upsilon_{1.1} + (k-2)^2] \quad (10)$$

Where

$\varrho_1 = \frac{kp_h}{2(kp_h+l)}$, $\varrho_2 = \frac{P}{2(P+\beta_2)}$, $\varrho_3 = \frac{P}{2(P+C_p)}$, $\varrho_4 = \frac{P}{2(P+\rho_p)}$, $\varrho_5 = \frac{\beta_2 P}{2(\beta_2 P+C_p)}$, $\varrho_6 = \frac{C_p P}{2(\beta_2+C_p P)}$,

$\varrho_7 = \frac{\beta_2 P}{2(\rho_p+C_p P)}$, $\varrho_8 = \frac{\rho_p P}{2(\rho_p P+C_p)}$, $\varrho_9 = \frac{\beta_2 P}{2(\beta_2 P+\rho_p)}$, $\varrho_{10} = \frac{\rho_p P}{2(\rho_p P+\beta_2)}$

$k = \frac{2\varrho_i^2\upsilon_{0.2}-2\varrho_i\upsilon_{1.1}+2}{2\varrho_i^2\upsilon_{0.2}-2\varrho_i\upsilon_{1.1}+2\upsilon_{2.0}+2}$.

Akintunde *et al.*[17] has been proposed the an improved estimation of population mean using auxiliary attribute in stratified sampling is

$$t_{5,j(st)} = \bar{y}_{st}\left[\alpha exp\left(\frac{(\omega p_{st}+L)-(\omega p_{st}+L)}{(\omega p_{st}+L)+(\omega p_{st}+L)}\right) + (1-\alpha)exp\left(\frac{(\omega p_{st}+L)-(\omega p_{st}+L)}{(\omega p_{st}+L)+(\omega p_{st}+L)}\right)\right] \quad (11)$$

The bias and MSE, respectively

$$Bias(t_{5,j(st)}) = \bar{Y}\left[\left(2\alpha-\frac{1}{2}\right)\zeta_i^2\upsilon_{0.2} - (2\alpha-1)\zeta_i\upsilon_{1.1}\right]$$

$$MSE(t_{5,j(st)}) = \bar{Y}^2[\upsilon_{2.0} + \zeta_j^2(4\alpha^2-4\alpha+1)\upsilon_{0.2} - 2\zeta_j(2\alpha-1)\upsilon_{1.1}] \quad (12)$$

Where ζ is an optimal constant that will be discovered later, in this case

$\zeta_1 = \frac{kp_h}{2(kp_h+l)}$, $\zeta_2 = \frac{P}{2(P+\beta_2)}$, $\zeta_3 = \frac{P}{2(P+C_p)}$, $\zeta_4 = \frac{P}{2(P+\rho_p)}$, $\zeta_5 = \frac{\beta_2 P}{2(\beta_2 P+C_p)}$, $\zeta_6 = \frac{C_p P}{2(\beta_2+C_p P)}$,

$\zeta_7 = \frac{\beta_2 P}{2(\rho_p+C_p P)}$, $\zeta_8 = \frac{\rho_p P}{2(\rho_p P+C_p)}$, $\zeta_9 = \frac{\beta_2 P}{2(\beta_2 P+\rho_p)}$, $\zeta_{10} = \frac{\rho_p P}{2(\rho_p P+\beta_2)}$

$\alpha = \frac{\upsilon_{1.1}}{2\zeta_j\upsilon_{0.2}} + \frac{1}{2}, j = 1,2,...,10.$

4. Proposed Estimator

This section evaluates the proposed estimator and examines its properties up to the first order of approximation when combining the ratio, product, exponential, and logarithmic types of estimators. Now the original version of the proposed estimator,

$$t_{pr(st)} = \bar{y}_{st}exp\left[\frac{\bar{P}-\bar{p}_{st}}{\bar{P}+\bar{p}_{st}}\right] + \bar{y}_{st}K_1 log\left[\frac{\bar{p}_{st}}{\bar{P}}\right] + K_2\bar{y}_{st}\left[1+log\frac{\bar{p}_{st}}{\bar{P}}\right] \quad (13)$$

Now put the values of $\bar{y}_{st}$ and P_{st} in (13) and simplify, we get

$$t_{pr(st)} = \bar{Y}\left[\left(1+e_0-\frac{1}{2}e_1-\frac{1}{2}e_0e_1+\frac{3}{8}e_1^2\right) + K_1\left(e_1+e_0e_1-\frac{1}{2}e_1^2\right) + K_2\left(1+e_0+e_1+e_0e_1-\frac{1}{2}e_1^2\right)\right]$$

Subtract $\bar{Y}$ on both sides, we get

$$t_{pr(st)} - \bar{Y} = \bar{Y}\left[\begin{array}{c}\left(e_0-\frac{1}{2}e_1-\frac{1}{2}e_0e_1+\frac{3}{8}e_1^2\right) + K_1\left(e_1+e_0e_1-\frac{1}{2}e_1^2\right)\\ + K_2\left(1+e_0+e_1+e_0e_1-\frac{1}{2}e_1^2\right)\end{array}\right] \quad (14)$$

The bias of the suggested estimator is given by taking expectations on both sides of (14)

$$Bias\left(t_{pr(st)}\right) = E\left[t_{pr(st)} - \bar{Y}\right]$$

$$E(t_{pr(st)} - \bar{Y}) = \bar{Y}E\begin{bmatrix}\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_1\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right) \\ + K_2\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\end{bmatrix}$$

$$E(t_{pr(st)} - \bar{Y}) = \bar{Y}\begin{bmatrix}E\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_1E\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right) \\ + K_2E\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\end{bmatrix}$$

The bias of the proposed estimator is given by,

$$Bias\left(t_{pr(st)}\right) = \bar{Y}\left[\left(\frac{3}{8}\upsilon_{0.2} - \frac{1}{2}\upsilon_{1.1}\right) + K_1\left(\upsilon_{1.1} - \frac{1}{2}\upsilon_{0.2}\right) + K_2\left(1 + \upsilon_{1.1} - \frac{1}{2}\upsilon_{0.2}\right)\right] \quad (15)$$

The MSE of the proposed estimator is now obtained by squaring and taking expectations on both sides of (14) as,

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

$$MSE\left(t_{pr(st)}\right) = \bar{Y}^2E\begin{bmatrix}\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_1\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right) \\ + K_2\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\end{bmatrix}^2$$

$$MSE\left(t_{pr(st)}\right) = \bar{Y}^2E\left(e_0^2 + \frac{1}{4}e_1^2 - e_0e_1\right) + K_1^2\bar{Y}^2E\left(e_1^2\right) + K_2^2\bar{Y}^2E\left(1 + 2e_0 + 2e_1 + e_0^2 + 4e_0e_1\right) + 2K_1\bar{Y}^2E\left(e_0e_1 - \frac{1}{2}e_1^2\right) + 2K_2\bar{Y}^2E\left(e_0 - \frac{1}{2}e_1 + e_0^2 - \frac{1}{8}e_1^2\right) + 2K_1K_2\bar{Y}^2E\left(e_0 + 2e_0e_1 - \frac{1}{2}e_1^2\right)$$

Now the MSE of proposed estimator is:

$$MSE\left(t_{pr(st)}\right) = \bar{Y}^2\left[A + K_1^2B + K_2^2C + 2K_1D + 2K_1K_2P + 2K_2Q\right] \quad (16)$$

The optimum values of K_1 and K_2 are given by,

$$K_1 = \frac{CD-PQ}{Q^2-BC}$$

$$K_2 = \frac{BP-DQ}{Q^2-BC}$$

Where,

$A = \upsilon_{2.0} - \upsilon_{1.1} + 0.25 * \upsilon_{0.2}$

$B = \upsilon_{0.2}$

$C = 1 + \upsilon_{2.0} + 4\upsilon_{1.1}$

$D = \upsilon_{1.1} - 0.5 * \upsilon_{0.2}$

$P = \upsilon_{2.0} - 0.125 * \upsilon_{0.2}$

$Q = 2\upsilon_{1.1} - 0.5 * \upsilon_{0.2}$.

4.1 Specific situations for the suggested estimator

In this section, stratified random sampling is used to explain the special situations of newly proposed and improved estimators for finite population. The first order of approximation's properties of the proposed estimator is explored.

4.1.1 Case-I

Put $K_2 = 0$ in proposed estimator (13) we get

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

Put the values of $\bar{y}_{st}$ and P_{st} in (17) and simplify, then we get

$$t'_{pr(st)} = \bar{Y}\left[\left(1 + e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{11}\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]$$

Subtract $\bar{Y}$ on both side, we get

$$t'_{pr(st)} - \bar{Y} = \bar{Y}\left[\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{11}\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right] \quad (18)$$

The bias of the suggested ratio-type estimator after taking into account expectations on both sides of (18) as,

$$Bias\left(t'_{pr(st)}\right) = E\left[t'_{pr(st)} - \bar{Y}\right]$$

$$E(t_{pr(st)} - \bar{Y}) = \bar{Y}E\left[\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{11}\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]$$

$$E(t_{pr(st)} - \bar{Y}) = \bar{Y}\left[E\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{11}E\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]$$

Then the bias in first case of proposed ratio-type estimator is obtained as

$$Bias\left(t_{pr(st)}\right) = \bar{Y}\left[\left(\frac{3}{8}\upsilon_{0.2} - \frac{1}{2}\upsilon_{1.1}\right) + K_{11}\left(\upsilon_{1.1} - \frac{1}{2}\upsilon_{0.2}\right)\right] \tag{19}$$

Now, in the first case of the suggested estimator, squaring and taking expectations on both sides of (18), we get the mean square error.

$$MSE\left(t'_{pr(st)}\right) = E[t'_{pr(st)} - \bar{Y}]^2$$

$$MSE\left(t'_{pr(st)}\right) = \bar{Y}^2E\left[E\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{11}E\left(e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]^2$$

$$MSE\left(t'_{pr(st)}\right) = \bar{Y}^2E\left(e_0^2 + \frac{1}{4}e_1^2 - e_0e_1\right) + K_{11}^2\bar{Y}^2E\left(e_1^2\right) + 2K_{11}\bar{Y}^2E\left(e_0e_1 - \frac{1}{2}e_1^2\right)$$

Then mean square error of proposed ratio-type estimator for first case is obtained as

$$MSE\left(t'_{pr(st)}\right) = \bar{Y}^2\left[A + K_{11}^2B + 2K_{11}D\right] \tag{20}$$

Where, K_{11} is the optimum value of proposed estimator for first case and given by

$K_{11} =- \left[\frac{D}{B}\right]$.

4.1.2 Case-II

Put $K_1 = 0$ in proposed estimator (3) we get

$$t''_{pr(st)} = \bar{y}_{st}exp\left[\frac{\bar{P}-p_{st}}{\bar{P}+\bar{p}_{st}}\right] + K_{12}\bar{y}_{st}log\left[1 + log\frac{\bar{p}_{st}}{\bar{P}}\right] \tag{21}$$

Put the values of $\bar{y}_{st}$ and P_{st} in (21) and simplify, then we get

$$t''_{pr(st)} = \bar{Y}\left[\left(1 + e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{12}\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]$$

Subtract $\bar{Y}$ on both side, we get

$$t''_{pr(st)} - \bar{Y} = \bar{Y}\left[\left(e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{12}\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right] \tag{22}$$

The bias of the proposed estimator, accounting for expectations on both sides of (22)

$$Bias\left(t''_{pr(st)}\right) = E\left[t''_{pr(st)} - \bar{Y}\right]$$

$$E(t''_{pr(st)} - \bar{Y}) = \bar{Y}E\left[\left(1 + e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{12}\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]$$

$$E(t''_{pr(st)} - \bar{Y}) = \bar{Y}\left[E\left(1 + e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{12}E\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]$$

Then the bias in second case of proposed estimator is obtained as

$$Bias\left(t''_{pr(st)}\right) = \bar{Y}\left[\left(\frac{3}{8}\upsilon_{0.2} - \frac{1}{2}\upsilon_{1.1}\right) + K_{12}\left(1 + \upsilon_{1.1} - \frac{1}{2}\upsilon_{0.2}\right)\right] \tag{23}$$

Now, square and take expectations on both sides of (22) and then estimate the suggested estimator's MSE in the second case

$$MSE\left(t''_{pr(st)}\right) = E[t''_{pr(st)} - \bar{Y}]^2$$

$$MSE\left(t''_{pr(st)}\right) = \bar{Y}^2E\left[\left(1 + e_0 - \frac{1}{2}e_1 - \frac{1}{2}e_0e_1 + \frac{3}{8}e_1^2\right) + K_{12}\left(1 + e_0 + e_1 + e_0e_1 - \frac{1}{2}e_1^2\right)\right]^2$$

$$MSE\left(t''_{pr(st)}\right) = \bar{Y}^2E\left(e_0^2 + \frac{1}{4}e_1^2 - e_0e_1\right) + K_{12}^2\bar{Y}^2E\left(1 + 2e_0 + 2e_1 + e_0^2 + 4e_0e_1\right) + 2K_{12}\bar{Y}^2E\left(e_0 - \frac{1}{2}e_1 + e_0^2 - \frac{1}{8}e_1^2\right)$$

Then MSE of second case of proposed estimator is obtained as

$$MSE\left(t''_{pr(st)}\right) = \bar{Y}^2\left[A + K_{12}^2 C + 2K_{12}P\right] \qquad (24)$$

Where, the second case of the suggested estimator's optimal value is K_{12} and is given by

$$K_{12} = -\left[\frac{P}{C}\right].$$

5. Empirical Study

In this section, we have considered two data sets for the verification of the theoretical findings.

Population 1: Auxiliary attribute: the number of apple trees greater than 15,000 in 1999; study variable: the amount of apples produced in 1999. (Source: Turkish Republic Institute of Statistics). The above data of Turkey have been divided into different regions (1: Marmara. 2: Agean 3: Mediterranean; 4: Central Anatolia; 5: The Black Sea and 6: East and Southeast Anatolia).

The samples whose sizes are calculated using the Neyman allocation method[20] were chosen at random from each stratum, see in **Table 1**.

Table 1. Data set

Population size	$N_1 = 106$	$N_2 = 106$	$N_3 = 94$	$N_4 = 171$	$N_5 = 204$	$N_6 = 173$
Sample size	$n_1 = 13$	$n_2 = 24$	$n_3 = 55$	$n_4 = 95$	$n_5 = 10$	$n_6 = 3$
Population mean	$\bar{Y}_1 = 1536.087$	$\bar{Y}_2 = 2212.54$	$\bar{Y}_3 = 9384.309$	$\bar{Y}_4 = 5588.012$	$\bar{Y}_5 = 966.956$	$\bar{Y}_6 = 404.399$
Sample mean	$P_1 = 0.245$	$P_2 = 0.292$	$P_3 = 0.468$	$P_4 = 0.485$	$P_5 = 0.363$	$P_6 = 0.116$
Population variance	$S_{y1} = 6425.087$	$S_{y2} = 11551.530$	$S_{y3} = 29907.48$	$S_{y4} = 28643.772$	$S_{y5} = 2389.771$	$S_{y6} = 945.748$
Auxiliary Population variance	$S_{P1} = 0.432$	$S_{P2} = 0.457$	$S_{P3} = 0.502$	$S_{P4} = 0.501$	$S_{P5} = 0.482$	$S_{P6} = 0.321$
Covariance	$S_{yP1} = 996.591$	$S_{yP2} = 1404.006$	$S_{yP3} = 48674.989$	$S_{yP4} = 2743.995$	$S_{yP5} = 449.021$	$S_{yP6} = 204.183$
Parameter	$\lambda_1 = 0.067$	$\lambda_2 = 0.032$	$\lambda_3 = 0.008$	$\lambda_4 = 0.005$	$\lambda_5 = 0.095$	$\lambda_6 = 0.328$
Weight	$W_1^2 = 0.015$	$W_2^2 = 0.015$	$W_3^2 = 0.012$	$W_4^2 = 0.040$	$W_5^2 = 0.057$	$W_6^2 = 0.041$
Parameter	$\beta_2 = -1.8$					

Population 2: One more attribute is the quantity of apples produced in 1998 that was less than 1000 tons. The study variable is the quantity of apples produced in 1999 (Source: Turkish Republic Institute of Statistics). By regions of Turkey, we have stratified the data (as 1: Marmara; 2: Agean; 3: Mediterranean; 4: Central Anatolia; 5: The Black Sea and 6: East and Southeast Anatolia). The samples from each stratum have been selected by the similar method as in Population 1. **Table 2** represents the summary of the data.

Table 2. Data set

Population size	$N_1 = 106$		$N_2 = 106$	$N_3 = 94$	$N_4 = 171$	$N_5 = 204$	$N_6 = 173$
Sample size	$n_1 = 13$		$n_2 = 24$	$n_3 = 55$	$n_4 = 95$	$n_5 = 10$	$n_6 = 3$
Population mean	$\bar{Y}_1 = 1536.087$		$\bar{Y}_2 = 2212.54$	$\bar{Y}_3 = 9384.309$	$\bar{Y}_4 = 5588.012$	$\bar{Y}_5 = 966.956$	$\bar{Y}_6 = 404.399$
Sample mean	$P_1 = 0.226$		$P_2 = 0.254$	$P_3 = 0.404$	$P_4 = 0.397$	$P_5 = 0.216$	$P_6 = 0.087$
Population variance	$S_{y1} = 6425.087$		$S_{y2} = 11551.530$	$S_{y3} = 29907.48$	$S_{y4} = 28643.772$	$S_{y5} = 2389.771$	$S_{y6} = 945.748$

Auxiliary Population variance	$S_{P1} = 0.421$		$S_{P2} = 0.4377$	$S_{P3} = 0.493$	$S_{P4} = 0.491$	$S_{P5} = 0.412$	$S_{P6} = 0.282$
Covariance	$S_{yP1} = 1029.204$		$S_{yP2} = 1481.372$	$S_{yP3} = 5408.425$	$S_{yP4} = 3190.691$	$S_{yP5} = 511.571$	$S_{yP6} = 190.918$
Parameter	$\lambda_1 = 0.067$		$\lambda_2 = 0.032$	$\lambda_3 = 0.008$	$\lambda_4 = 0.005$	$\lambda_5 = 0.095$	$\lambda_6 = 0.328$
Weight	$W_1^2 = 0.015$		$W_2^2 = 0.015$	$W_3^2 = 0.012$	$W_4^2 = 0.040$	$W_5^2 = 0.057$	$W_6^2 = 0.041$
Parameter	$\beta_2 =- 1.8$						

The PRE for both the considered populations have been presented in **Table 3**.

Table 3. PRE of different estimators for Population I and Population II

Estimators	Population-I PRE	Population-II PRE
$t_{1(st)}$	100.00000	100.00000
$t_{2(st)}$	131.64145	189.83977
$t_{3(st)}$	145.74597	190.01602
$t_{4,1(st)}$	149.47386	146.32628
$t_{4,2(st)}$	121.15516	157.38293
$t_{4,3(st)}$	138.01719	179.27133
$t_{4,4(st)}$	147.37997	186.67924
$t_{4,5(st)}$	91.84414	117.07463
$t_{4,6(st)}$	118.12120	153.10773
$t_{4,7(st)}$	147.96787	186.67459
$t_{4,8(st)}$	132.41424	172.56485
$t_{4,9(st)}$	146.18799	173.16195
$t_{4,10(st)}$	128.98899	168.11630
$t_{5,1(st)}$	131.64145	156.00983
$t_{5,2(st)}$	102.76677	133.15559
$t_{5,3(st)}$	122.55306	158.88311
$t_{5,4(st)}$	131.73787	164.42912
$t_{5,5(st)}$	61.64805	73.83767
$t_{5,6(st)}$	98.96330	127.69696
$t_{5,7(st)}$	132.13673	151.48865
$t_{5,8(st)}$	116.25518	136.65959
$t_{5,9(st)}$	123.59916	146.30112
$t_{5,10(st)}$	112.25868	135.63916
$t'_{pr(st)}$	**398.75098**	**523.86242**
$t''_{pr(st)}$	**395.29143**	**517.65141**
$\boldsymbol{t_{pr(st)}}$	**399.92377**	**525.38788**

The above results have also been presented in the form of graph in **Figure 1**.

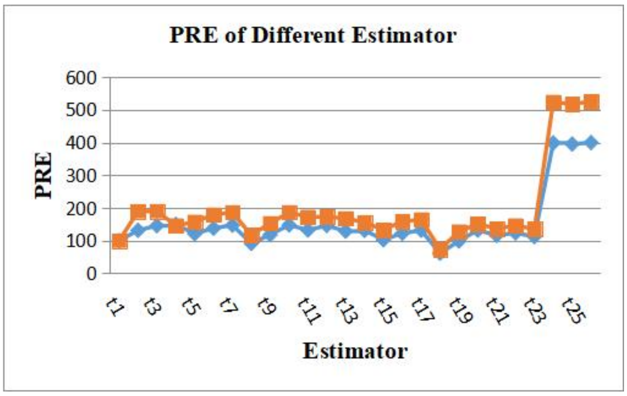

Figure 1. PRE of various estimators with respect to mean estimator for Population-I and II.

Here the estimators $t_{1(st)}, t_{2(st)}, t_{3(st)}, t_{4j(st)}, t_{5j(st)}$, $t'_{pr(st)}, t''_{pr(st)}$ and $t_{pr(st)}$ are presented from left to right in **Figure 1**.

6. Results and Discussion

In this paper, some members of the existing estimators are discussed and the bias and MSE of these estimators are obtained for an approximation of order one. Using two real population data sets, population 1 in **Table 1** and population 2 in **Table 2** are given. It may be observed from **Table 3** that the efficiency of the proposed estimator is the highest is **399.92377** and the PRE of the estimators lie in the interval [61.64805 to 398.75098] for population I. From **Table 3**, it is evident that the efficiency of the proposed estimator is the highest and is **525.38788** while the PRE of the estimators with respect to mean estimator lie in the interval [73.83767 to 523.86242] for population II.

7. Conclusion

This paper introduces an improved ratio-product-cum-logarithmic type estimator that estimate the population mean using auxiliary attribute by implying one auxiliary variable in stratified random sampling and find MSE and PRE. It has also been demonstrated theoretically that the proposed estimator is more efficient than other estimators already in use. The numerical example confirms that the theoretical results can be applied in practice. From **Table 3**, it has been observed that the proposed estimator has the least MSE, and highest PRE from **Table 3**. Thus, the proposed estimator may be utilized for a more efficient estimation of the population mean in various application fields, including agricultural sciences, biological sciences, commerce, economics, engineering sciences, medical sciences, social sciences and others areas of applications.

Acknowledgement

The authors are thankful to the editor and the learned referees for their critical observations which improved the earlier draft.

References

1. Cochran W G. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. The Journal of Agricultural Science 1940; 30(2): 262-275.
2. Srivastava S K. A class of estimators using auxiliary information in sample surveys. Canadian Journal of Statistics 1980; 8(2): 253-254.
3. Sisodia J, Scientist B V S, Dwivedi J V K. A class of ratio-cum-product-type estimator. Biometrical Journal 1981; 23(2): 133-139.
4. Bahl S, Tuteja R K. Ratio and product type estimator. Journal of Information and Optimization Sciences 1991; 9: 159-163.
5. Singh P, Vishwakarma K. Modified exponential ratio and product estimators for finite population mean in double sampling. Australian Journal of Statistics 2007; 36: 217-225.
6. Koyuncu N. Families of estimators for population mean using in-formation on auxiliary attribute in stratified random sampling. Gazi University Journal of Science 2013; 26(2): 181-193.
7. Haq A, Shabbir J. Improved family of ratio estimators in simple and stratified random sampling. Communications in Statistics-Theory and Methods 2013; 42(5): 782-799.
8. Sanaullah A, Ali H A, ul Amin M N, *et al.* Generalized exponential chain ratio estimators under stratified two phase random sampling. Applied Mathematics and Computation 2014; 226: 541-547.
9. Shabbir J, Gupta S. On generalized exponential chain ratio estimators under stratified two-phase random sampling. Communications in Statistics-Theory and Methods 2017; 46(6): 2910-2920.
10. Lone H A, Tailor R. Estimation of population variance in simple random sampling. Journal of Statistics and Management Systems 2017; 20(1): 17-38.
11. Cekim H O, Kadilar C. New families of unbiased estimators in stratified random sampling. Journal of Statistics and Management Systems 2018; 21(8): 1481-1499.
12. Yadav S K, Mishra S S, Gupta S. An efficient estimator for population variance using parameters of an auxiliary variable. Journal of Statistics and Management Systems 2019; 22(6): 1005-1013.
13. Zaman T. Efficient estimators of population mean using auxiliary attribute in stratified random sampling. Advances and Applications in Statistics 2019; 56(2): 153-171.
14. Zaman T, Kadilar C. On estimating the population mean using auxiliary character in stratified random sampling. Journal of Statistics and Management Systems 2020; 23(8): 1415-1426.
15. Yadav S K, Vishwakarma G K, Varshney R, *et al.* Improved memory type product estimator for population mean in stratified random sampling under linear cost function. SN Computer Science 2023; 4(3): 235.
16. Yadav S K, Arya D, Vishwakarma G K, *et al.* Generalized ratio-cum-exponential-log ratio type estimators of population mean under simple random sampling scheme. Lobachevskii Journal of Mathematics 2023; 44(9): 3889-3901.
17. Akintunde A A, Olayiwola O M, Adewole A I, *et al.* An improved estimation of population mean using auxiliary attribute in stratified sampling. Nigerian Journal of Pure Applied Sciences 2023; 36(1): 4463-4471.
18. Yadav S K, Kumar Verma M, Varshney R. Optimal strategy for elevated estimation of population mean in stratified random sampling under linear cost function. Annals of Data Science 2024; 11(2). https://doi.org/10.1007/s40745-024-00520-9
19. Koyuncu N, Kadilar C. Efficient estimators for the population mean. Hacettepe Journal of Mathematics and Statistics 2009; 38(2): 217-225.
20. Mathew O O, Sola A F, Oladiran B H, *et al.* Efficiency of Neyman allocation procedure over other allocation procedures in stratified random sampling. American Journal of Theoretical and Applied Statistics 2013; 2(5): 122-127.
21. Singh H P, Solanki R S. Improved estimation of population mean in simple random sampling using information on auxiliary attribute. Applied Mathematics and Computation 2012;

218(15): 7798-7812.

22. Singh H P, Gupta A, Tailor R. Efficient class of estimators for finite population mean using auxiliary attribute in stratified random sampling. Scientific Reports 2023; 13(1): 10253.
23. Koyuncu N. Families of estimators for population mean using information on auxiliary attribute in stratified random sampling. Gazi University Journal of Science 2013; 26(2): 181-193.
24. Ahmadini A A H, Yadav T, Yadav S K, *et al*. Restructured searls family of estimators of population mean in the presence of non-response. Frontiers in Applied Mathematics and Statistics 2022; 8: 969068.