Method Article

Optimum estimator in simple random sampling using two auxiliary attributes with application in agriculture, fisheries and education sectors

Monika Saini, Bhatt Ravi Jitendrakumar, Ashish Kumar*
Department of Mathematics & Statistics, Manipal University Jaipur, Jaipur 303007, India

ABSTRACT

In modern age of information technology, data is available everywhere in huge amount. Every sector generates lot of data every day. The investigation of each unit of data is not feasible due to limited resources like time, labor, and cost. In such situations, survey sampling is recommended to draw the information about the population parameters. Therefore, the main objective of present study is to develop an estimation method for obtaining the information about population parameter. We propose an optimum estimator for enhanced estimation of population mean in simple random sampling by utilizing the information of the two auxiliary attribute. The expression for bias, mean squared error (MSE) and minimum mean squared error of the proposed estimator are derived up to the first order of approximation and it is shown that the proposed estimator under derived conditions perform better than the existing estimators theoretically. Four population are demonstrated to assess the performance as well as applicability of the proposed estimator. The percentage relative efficiency (PRE) of proposed estimator for all the populations is 209.533, 163.852, 210.398 and 340.578, respectively. The numerical illustrations confirm that the proposed estimator dominates over the existing estimators.

- The main objective of present study is to propose a new estimator/method for estimation of population mean using two auxiliary attributes under simple random sampling.
- The bias and mean square error of the proposed estimator/method is derived and compared with the existing estimators to compare the efficiency theoretically.
- Applications of the proposed method/estimator is highlighted using thorough the real data sets of various sectors.

© 2022 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

ARTICLE INFO
Method name: Optimum estimator for estimating population mean using two auxiliary attributes
Keywords: Finite population, Auxiliary attribute, Simple random sampling, Bias, Mean squared error
Article history: Received 13 September 2022; Accepted 31 October 2022; Available online 9 November 2022

* Corresponding author.
E-mail address: ashish.kumar@jaipur.manipal.edu (A. Kumar).

https://doi.org/10.1016/j.mex.2022.101915
2215-0161/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Specifications table

| Subject Area: | Mathematical Statistics |
|---------------|-------------------------|
| More specific subject area: | Sample Surveys |
| Method name: | Optimum estimator for estimating population mean using two auxiliary attributes |
| Name and references of original method: | The proposed method is motivated by following references: |

- Ahmad, S., Arslan, M., Khan, A., & Shabbir, J. (2021). A generalized exponential-type estimator for population mean using auxiliary attributes. *Plos one*, 16(5), e0246947.
- Zaman, T., & Kadilar, C. (2019). Novel family of exponential estimators using information of auxiliary attribute. *Journal of Statistics and Management Systems*, 22(8), 1499–1509
- Lu, J. (2017). Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural, biomedical, and power engineering. *Mathematical Problems in Engineering*.

| Resource availability: | Data utilized in the analysis is available in public domain. |

Background

In sample surveys, it is well documented in Cochran [20] that the use of supplementary information provided by auxiliary variables or attributes is frequently used for increasing the precision of the estimators by taking the advantages of correlation between the study variable and auxiliary variable. Regression, ratio, and product estimators are good examples in this perspective. Though Cochran [1] investigated that ratio estimator is best suited when the study and auxiliary variable are highly positively correlated whereas in case of highly negative correlated variables, product method of estimation is better. In many real-life situations, the study variable is not always quantitative in nature. The responses recorded from respondents are qualitative in such situations the recorded information is called attributes. Several studies like Shabbir and Gupta [22] and Abd-Elfattah et al. [23] have been conducted to improve the precision of the estimator by utilizing the auxiliary attributes having applications in agriculture, health science, fisheries, power engineering etc. It is also investigated that use of more than one auxiliary attribute enhance the efficiency of the estimator. Singh et al. [2] used auxiliary attribute information for establishment of ration estimators in simple random sampling (SRS). The bias and mean square error of the estimator has been computed for existing data set available in literature. It is developed as a modified estimator of Koyuncu and Kadilar [21] estimator and proved that it outperformed the existing estimators. Singh and Kumar [4] used auxiliary information to estimate improved regression estimator for SRS. Here the auxiliary information is qualitative, and concept of non-response incorporated in estimation. Malik and Singh [3] initiated the use of two auxiliary variables in estimation of population mean and proposed enhanced estimators. Here, auxiliary information available in qualitative form and this estimator performed better than simple regression estimator. Ekpenyong and Enang [5] suggested better exponential estimators in SRS for estimating population mean. The concept of simple random sampling without replacement used for development of estimator. Lu [6] explored the applications of estimators developed under auxiliary information in agriculture and power engineering sectors. The estimator compared with regression estimator and other existing estimators and proved efficient. Zaman and Kadilar [7] utilized auxiliary information for development of a novel family of exponential estimators. Ahmad et al. [8] carried out the generalization of exponential ratio estimators under auxiliary estimators. The estimator was exponential-based estimator while estimator proposed by us is mixed type estimator while estimator proposed in present study is a mixture of simple, ratio and product estimators. Mahajan et al. [9] explored the applications of estimation and sample surveys in agriculture and health sciences. Kumar and Saini [10] suggested a predictive approach to estimate the population mean under auxiliary attribute. Yunusa et al. [11] utilized auxiliary variables in development of regression type estimators to estimation the population mean. Rather et al. [24] used auxiliary information for development of a mixed exponential ratio type estimator for estimating the population mean. The simple random sampling and double sampling techniques utilized for selection
of the sample. Zaman et al. [25] proposed an exponential type estimator for assessing and estimating the COVID-19 risk in various countries. Two multivariate families of exponential type estimators proposed by utilizing the information on two auxiliary variables. Many authors like Wayangkau et al. [18], Waheeb et al. [15], Rajak [16], and Jabal et al. [17] discussed some other data analysis techniques for analysis of agricultural information. The above cited literature motivates to explore the applicability of two auxiliary attributes in estimation of mean of various populations associated with agriculture, fisheries, and education sectors. The primary goal of this paper is to propose a novel optimum estimator for estimating finite population mean using auxiliary attributes. The expressions for the bias and mean square error (MSE) of the proposed estimator are inferred up to the first order of approximation. On the bases of theoretical and numerical comparisons, we demonstrate that the proposed estimator is more efficient than existing estimators.

Material and methods

Consider \( \gamma = (\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n) \) be a finite population of size \( N \). We draw a sample of size \( n \) (with \( n < N \)) units from \( \gamma \) using simple random sample without replacement (SRSWOR). Let \( \gamma_i \) be the study variable and \( \delta_i \) be the characteristics of the auxiliary attributes i.e. \( \delta_i = 1 \) if the \( i \)th unit possess attribute and \( \delta_i = 0 \), otherwise. Let \( g = \sum_{i=1}^{N} \delta_i \) be the total number of units in the population possessing attributes \( \delta_i \) and \( g = \sum_{i=1}^{N} \delta_i \) be the total number of units in the sample possessing attributes \( \delta_i \). Let \( D = (G/N) \) be the proportion of units in the population and \( d = (\bar{g}/n) \) be the population of units in the sample. Let \( \bar{Y} = \frac{\sum_{i=1}^{N} \gamma_i}{N} \) and \( \bar{y} = \frac{\sum_{i=1}^{n} \gamma_i}{n} \) be the population and sample mean of the study variable. Let \( D_1 \) and \( D_2 \) be the population proportion of the auxiliary attributes and sample proportion of auxiliary attributes is denoted by \( d_1 \) and \( d_2 \). Let \( S^2 = \frac{\sum_{i=1}^{N} (\gamma_i - \bar{Y})^2}{N-1} \) be the population variance of the study variable \( y \). Let \( S^2_1 = \frac{\sum_{i=1}^{N} (d_1 - D_1)^2}{N-1} \) and \( S^2_2 = \frac{\sum_{i=1}^{N} (d_2 - D_2)^2}{N-1} \) respectively be the population variance of the auxiliary attributes \( d_1 \) and \( d_2 \).

Let \( \beta_j = \frac{S^2_{yj}}{S^2_j} \) be the coefficient of variation of the auxiliary \( d_1 \) and \( d_2 \). Let \( S_{ydj} = \frac{\sum_{i=1}^{N} (\gamma_i - \bar{Y})(d_j - D_j)}{N-1} \) be the population covariance between the study variable \( y \) and the auxiliary attributes \( d_j (j = 1, 2) \). Let \( S_{d1d2} = \frac{\sum_{i=1}^{N} (d_1 - D_1)(d_2 - D_2)}{N-1} \) be the population covariance between the auxiliary attributes \( d_1 \) and \( d_2 \). Let \( \rho_{ydj} = \frac{S_{ydj}}{\sqrt{S^2_{yj}S^2_j}} \) be the population point bi-serial correlation coefficient between the study variable \( y \) and the auxiliary attribute \( d_j (j = 1, 2) \). Let \( \pi_0 = \bar{y} \); \( \pi_1 = \frac{d_1 - D_1}{D_1} \) and \( \pi_2 = \frac{d_2 - D_2}{D_2} \) be the error terms such that \( E[\pi_0] = 0 \) (i = 0, 1, 2), \( E[\pi_1^2] = \xi C^2_{\pi_1} \), \( E[\pi_2^2] = \xi C^2_{\pi_2} \), \( E[\pi_1\pi_2] = \xi C_{\pi_1\pi_2} \) and \( \xi = \frac{1}{N} - \frac{1}{n} \) and \( f = \frac{n}{N} \).

The stepwise framework of proposed method is described as follows:

Step 1: Consider a finite population of size \( N \).
Step 2: Select a random sample of size \( n \) from the population using simple random sampling without replacement.
Step 3: Observe \( y_i \) and \( \delta_i \) from sampling units.
Step 4: Define the expressions for population and sample characteristics.
Step 5: Propose the estimator for estimating population mean using two auxiliary attributes and derive its properties.
Step 6: Compare the proposed estimator with the existing estimators theoretically and numerically.

Existing estimators

Unbiased estimator \( (\hat{\mu}_0) \)

The most widely used estimator discussed by Cochran [1] of population mean \( \bar{Y} \) of the study variable, is given by

\[
\hat{\mu}_{0} = \bar{y}.
\]
sample mean is an unbiased estimator of population mean and upto the first order of approximation the variance of $\mathcal{M}$ is given by
\[ V(\hat{\mu}_0) = V(\bar{Y}) = \xi \hat{Y}^2 \sigma_2^2. \] (2)

\textbf{Naik and Gupta ($\hat{\mu}_1$)}

Naik and Gupta [12] proposed the following ratio estimator of population mean $\bar{Y}$ when the population proportion $D_1$ of auxiliary attribute is known
\[ \hat{\mu}_1 = \bar{Y} \left( \frac{D_1}{d_1} \right). \] (3)

The bias and MSE of $\hat{\mu}_1$ to the first order of approximation is given by,
\[ \text{Bias}(\hat{\mu}_1) = \xi \bar{Y} \left[ \rho_{yd1} c d_2 - \rho_{yd1} Y c d_2 \right]. \]
and
\[ \text{MSE}(\hat{\mu}_1) = \xi \bar{Y}^2 \left[ c_2^2 + c_2^2 - 2 \rho_{yd1} Y c d_2 \right]. \] (4)

\textbf{Naik and Gupta ($\hat{\mu}_2$)}

Naik and Gupta [12] proposed the following product estimator of population mean $\bar{Y}$ when the population proportion $D_1$ of auxiliary attribute is known
\[ \hat{\mu}_2 = \bar{Y} \left( \frac{d_1}{D_1} \right). \] (5)

The bias and MSE of $\hat{\mu}_2$, to the first order of approximation is given by
\[ \text{Bias}(\hat{\mu}_2) = \bar{Y} \xi \left[ \rho_{yd1} c d_2 \right]. \]
and
\[ \text{MSE}(\hat{\mu}_2) = \bar{Y}^2 \left[ c_2^2 + c_2^2 + 2 \rho_{yd1} Y c d_2 \right]. \] (6)

\textbf{Singh et al. ($\hat{\mu}_3$)}

Singh et al. [2] suggested an exponential type ratio estimator for estimating population mean $\bar{Y}$ using the population proportion $D_1$ of auxiliary attribute is known
\[ \hat{\mu}_3 = \bar{Y} \exp \left[ \frac{D_1 - d_1}{D_1 + d_1} \right]. \] (7)

The bias and MSE of $\hat{\mu}_3$, to the first order of approximation,
\[ \text{Bias}(\hat{\mu}_3) = \bar{Y} \xi \left[ \frac{3}{8} c_2^2 - \frac{1}{2} \rho_{yd1} c d_2 \right]. \]
and
\[ \text{MSE}(\hat{\mu}_3) = \bar{Y}^2 \left[ c_2^2 + \frac{1}{4} c_2^2 - \rho_{yd1} Y c d_2 \right]. \] (8)

\textbf{Singh et al. ($\hat{\mu}_4$)}

Singh et al. [2] suggested the following product estimators for estimating population mean $\bar{Y}$ when the population proportion $D_1$ of auxiliary attribute is known
\[ \hat{\mu}_4 = \bar{Y} \exp \left[ \frac{d_1 - D_1}{d_1 + D_1} \right]. \] (9)

Similarly, the bias and MSE of $\hat{\mu}_4$, is given by
\[ \text{Bias}(\hat{\mu}_4) = \frac{1}{2} \bar{Y} \xi \left[ \rho_{yd1} c d_2 - \frac{1}{4} c_2^2 \right]. \]
and
\[ \text{MSE}(\hat{\mu}_4) = \frac{1}{2} \bar{Y}^2 \xi \left[ c_2^2 + \frac{1}{4} c_2^2 + \rho_{yd1} Y c d_2 \right]. \] (10)
Kumar and Bhougal ($\hat{\mu}_5$)

Kumar and Bhougal [13] proposed an exponential type of ratio-product estimator for estimating population mean $\bar{Y}$ when the population proportion $D_1$ of auxiliary attribute is known

$$\hat{\mu}_5 = \bar{y} \left[ \alpha \exp \left( \frac{D_1 - d_1}{D_1 + d_1} \right) + (1 - \alpha) \exp \left( \frac{D_1 - d_1}{D_1 + d_1} \right) \right],$$  \hspace{1cm} (11)

where $\alpha$ is unknown constant.

The bias and MSE of $\hat{\mu}_5$ to the first order of approximation is given by

$$Bias(\hat{\mu}_5) \approx \bar{Y} \bar{\xi} \left[ \frac{1}{8} (4\alpha - 1) c_d^2 - \left( \alpha - \frac{1}{2} \right) \rho_{yd1} c_y c_d1 \right],$$

and

$$MSE(\hat{\mu}_5) \approx \bar{Y} \bar{\xi}^2 c_y^2 \left[ 1 - \rho_{yd1}^2 \right].$$  \hspace{1cm} (12)

The optimum value of $\alpha$ is given by

$$\alpha_{opt} \approx \frac{1}{2} + \rho_{yd1} c_y / c_p1.$$

Singh and Kumar ($\hat{\mu}_6$)

Singh and Kumar [4] suggested ratio estimator for estimating population mean $\bar{Y}$ when the population proportion $D_1$ and $D_2$ of auxiliary attribute are known

$$\hat{\mu}_6 = \bar{y} \left( \frac{D_1}{d_1} \right) \left( \frac{D_2}{d_2} \right).$$  \hspace{1cm} (13)

The bias and MSE of $\hat{\mu}_6$ to the first order of approximation are given by

$$Bias(\hat{\mu}_6) \approx \bar{Y} \bar{\xi} \left[ c_d + c_d2 + \rho_{d1d2} c_{d1} c_{d2} - \rho_{yd1} c_y c_{d1} - \rho_{yd2} c_y c_{d2} \right],$$

and

$$MSE(\hat{\mu}_6) \approx \bar{Y} \bar{\xi}^2 \left[ c_y^2 + c_{d1}^2 + c_{d2}^2 - 2 \left( \rho_{yd1} c_y c_{d1} - \rho_{yd1} c_y c_{d1} - \rho_{yd2} c_y c_{d2} \right) \right].$$  \hspace{1cm} (14)

Singh and Kumar ($\hat{\mu}_7$)

Singh and Kumar [4] suggested product estimator for estimating population mean $\bar{Y}$ when the population proportion $D_1$ and $D_2$ of auxiliary attribute are known

$$\hat{\mu}_7 = \bar{y} \left( \frac{d_1}{D_1} \right) \left( \frac{d_2}{D_2} \right).$$  \hspace{1cm} (15)

The bias and MSE of $\hat{\mu}_7$ to the first order of approximation are given by

$$Bias(\hat{\mu}_7) \approx \bar{Y} \bar{\xi} \left[ \rho_{d1d2} c_{d1} c_{d2} - \rho_{yd1} c_y c_{d1} + \rho_{yd2} c_y c_{d2} \right],$$

$$MSE(\hat{\mu}_7) \approx \bar{Y} \bar{\xi}^2 \left[ c_y^2 + c_{d1}^2 + c_{d2}^2 + 2 \left( \rho_{yd1} c_y c_{d1} - \rho_{yd1} c_y c_{d1} + \rho_{yd2} c_y c_{d2} \right) \right].$$  \hspace{1cm} (16)

Ahmed et al. ($\hat{\mu}_8$)

Ahmed et al. [8] proposed a generalized class of factor type of estimators for estimating population mean $\bar{Y}$ when the population proportion $D_1$ and $D_2$ of auxiliary attribute are known is given by

$$\hat{\mu}_8 = \bar{y} \left[ \exp \left( S_1 + M_1 \right) \exp \left( S_2 + M_2 \right) \right]$$  \hspace{1cm} (17)

where,

$$S_1 = (A_1 + C_1)D_1 + JB_1 d_1, \quad S_2 = (A_2 + C_2)D_2 + JB_2 d_2, \quad M_1 = (A_1 +JB_1)D_1 + C_1 d_1, \quad M_2 = (A_2 + JB_2)D_2 + C_2 d_2, \quad A_i = (K_i - 1)(K_i - 2), \quad B_i = (K_i - 1)(K_i - 4), \quad C_i = (K_i - 2)(K_i - 3)(K_i - 4).$$
The bias and MSE of $\hat{\mu}_g$ to the first order of approximation are given by

$$\text{Bias}(\hat{\mu}_g) = \bar{y} \xi \left[ \frac{1}{2} \sigma_1 \rho_{yd1} C_{d1} + \frac{1}{2} \sigma_2 \rho_{yd2} C_{d2} + \frac{1}{8} C_{d1}^2 (\sigma_1^2 - 2\sigma_1 v_1) + \frac{1}{8} C_{d2}^2 (\sigma_2^2 - 2\sigma_2 v_2) \right].$$

where $v_1 = \frac{\rho_{B1} + C_1}{A_1 + \rho_{B1} + C_1}$ and $v_2 = \frac{\rho_{B2} + C_2}{A_2 + \rho_{B2} + C_2}$.

The minimum MSE of $(\hat{\mu}_g)$ to the first order of approximation is given by

$$\text{MSE}(\hat{\mu}_g) \equiv \varphi^2 C_g^2 \left[ 1 - R^2_{yd1d2} \right].$$

where

$$R^2_{yd1d2} = \frac{\rho^2_{yd1} + \rho^2_{yd2} - 2\rho_{yd1} \rho_{yd2}}{1 - \rho^2_{yd1d2}}.$$

The optimum value of $\sigma_{10pt}$ and $\sigma_{20pt}$ are

$$\sigma_{10pt} = \frac{2C_y (\rho_{yd1} - \rho_{d1d2} \rho_{yd2})}{C_{d1} (\rho^2_{d1d2} - 1)} \quad \text{and} \quad \sigma_{20pt} = \frac{2C_y (\rho_{yd2} - \rho_{d1d2} \rho_{yd1})}{C_{d2} (\rho^2_{d1d2} - 1)}.$$

**Proposed estimator & its properties**

Motivated by Singh and Espejo [14] and Lu [6], adopting the same procedure we proposed as estimator for estimating population mean by utilizing two auxiliary attributes as

$$\left( \hat{\mu}_{pop} \right) = \frac{\omega_1 \bar{y} + \omega_2 (D_1 - d_1) + \omega_3 (D_2 - d_2)}{4} \left( \frac{D_1}{d_1} + \frac{d_1}{D_1} \right) \left( \frac{D_2}{d_2} + \frac{d_2}{D_2} \right)$$

To obtain the bias and MSE of proposed estimator, applying the error approximation into Eq. (17), in terms of $\pi$’s, the proposed estimator can be expressed as

$$\hat{\mu}_{prop} = \frac{\omega_1 \bar{y} (\pi_0 + 1) - \omega_2 D_1 \pi_1 - \omega_3 D_2 \pi_2}{4} \left( \frac{1}{\pi_1 + 1} + \pi_1 + 1 \right) \left( \frac{1}{\pi_2 + 1} + \pi_2 + 1 \right)$$

Expanding the right-hand side of [20], we get

$$\hat{\mu}_{prop} = \frac{\omega_1 \bar{y} (\pi_0 + 1) - \omega_2 D_1 \pi_1 - \omega_3 D_2 \pi_2}{4} \left[ \pi_1 + 1 + (1 - \pi_1 + \pi_1^2 + \ldots) \right] \left[ \pi_2 + 1 + (1 - \pi_2 + \pi_2^2 + \ldots) \right].$$

In Eq. (21) we will neglect the terms of $\pi$’s, power having greater than two, we get

$$\hat{\mu}_{prop} - \bar{y} \cong (\omega_1 - 1) \bar{y} + \omega_1 \bar{y} \pi_0 - \omega_2 D_1 \pi_1 - \omega_3 D_2 \pi_2 + \frac{1}{2} \omega_1 \bar{y} \pi_1^2 + \frac{1}{2} \omega_1 \bar{y} \pi_2^2.$$

To get the bias of the proposed estimator, we need to take expectation on both the sides of (22), hence we will get the bias of the proposed estimator up to the first order of approximation (Fig. 1).

$$E(\hat{\mu}_{prop} - \bar{y}) \cong (\omega_1 - 1) \bar{y} + \omega_1 \bar{y} E(\pi_0) - \omega_2 D_1 E(\pi_1) - \omega_3 D_2 E(\pi_2) + \frac{1}{2} \omega_1 \bar{y} E(\pi_1^2) + \frac{1}{2} \omega_1 \bar{y} E(\pi_2^2).$$

The bias of the proposed estimator is given by

$$\text{Bias}(\hat{\mu}_{prop}) = E(\hat{\mu}_{prop} - \bar{y}) \equiv \bar{y} \left( (\omega_1 - 1) + \frac{1}{2} \xi_1 C_{d1}^2 + \frac{1}{2} \xi_1 C_{d2}^2 \right).$$
Now for deriving the MSE of the proposed estimator, let us square both sides of (22), neglecting terms \( \pi \)'s having power greater than two, and taking expectation on both sides, after simplification we get,

\[
\text{MSE}(\hat{\mu}_{\text{prop}}) = E(\hat{\mu}_{\text{prop}} - \bar{Y})^2. \\
\equiv \bar{Y}^2\left((\omega_1 - 1)^2 - \tilde{\xi}\omega_1(C_{d1}^2 + C_{d2}^2) + \tilde{\xi}\omega_1(C_{d1}^2 + C_{d2}^2 + C_{d1}^2) - 2\tilde{\xi}\omega_1\bar{Y}C_\omega(\omega_2C_{d1}\rho_{yd1}D_1 + \omega_3C_{d2}\rho_{yd2}D_2) \\
+ \tilde{\xi}\left(\omega_2C_{d1}D_1^2 + 2C_{d1}C_{d2}\omega_2\rho_{yd1}D_1D_2 + \omega_3C_{d2}^2D_2^2\right)\right].
\]  
(23)

To obtain the optimum value of \( \omega_1, \omega_2 \) and \( \omega_3 \), we partially differentiate the Eq. (23) with respect to \( \omega_1, \omega_2 \) and \( \omega_3 \) and put it equal to zero, the optimum values of \( \omega_1, \omega_2 \) and \( \omega_3 \) are given by

\[
\omega_1^* = \frac{\beta\left[2 + \tilde{\xi}\left(C_{d1}^2 + C_{d2}^2\right)\right]}{2\left(\beta + C_{d1}^2\beta\tilde{\xi} + \tilde{\xi}\left(C_{d1}^2 + C_{d2}^2\right)\right)},
\]

\[
\omega_2^* = C_\rho\bar{Y}(\rho_{yd1} - \rho_{yd2}\rho_{yd2})\left[2 + \tilde{\xi}\left(C_{d1}^2 + C_{d2}^2\right)\right],
\]

\[
\omega_3^* = C_\rho\bar{Y}(\rho_{yd2} - \rho_{yd1}\rho_{yd1})\left[2 + \tilde{\xi}\left(C_{d1}^2 + C_{d2}^2\right)\right],
\]

The minimum MSE of \( \hat{\mu}_{\text{prop}} \) can be shown as,

\[
\text{MSE}(\hat{\mu}_{\text{prop}})_{\min} = \frac{\xi\bar{Y}^2\left[4C_{d1}^2\alpha - \beta\tilde{\xi}\left(C_{d1}^2 + C_{d2}^2\right)\right]}{4\left(\beta + C_{d2}^2\beta\tilde{\xi} + \tilde{\xi}\left(C_{d1}^2 + C_{d2}^2\right)\right)},
\]  
(24)

where \( \alpha = 1 - \rho_{d1d2}^2 - \rho_{yd1}^2 + 2\rho_{yd1}\rho_{yd2}\rho_{yd2} - \rho_{yd2}^2 \) and \( \beta = 1 - \rho_{d1d2}^2 \).

Theoretical comparison of estimators

In this section we compare our proposed estimator with the exiting estimators discussed in Section 2, we will have the conditions as follows:
From Eqs. (2) and (24)
\[
\text{MSE}(\hat{\mu}_0) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
c_y^2 > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (4) and (24)
\[
\text{MSE}(\hat{\mu}_1) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
\left[c_y^2 + c_{d1}^2 - 2\rho_{yd1}c_yc_{d1}\right] > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (6) and (24)
\[
\text{MSE}(\hat{\mu}_2) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
\left[c_y^2 + c_{d1}^2 + 2\rho_{yd1}c_yc_{d1}\right] > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (8) and (24)
\[
\text{MSE}(\hat{\mu}_3) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
\left[c_y^2 + \frac{1}{4}c_{d1}^2 - \rho_{yd1}c_yc_{d1}\right] > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (10) and (24)
\[
\text{MSE}(\hat{\mu}_4) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
\left[c_y^2 + \frac{1}{4}c_{d1}^2 + \rho_{yd1}c_yc_{d1}\right] > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (12) and (24)
\[
\text{MSE}(\hat{\mu}_5) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
c_y^2[1 - \rho_{yd1}^2] > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (14) and (24)
\[
\text{MSE}(\hat{\mu}_6) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
\left[c_y^2 + c_{d1}^2 + c_{d2}^2 - 2\left(\rho_{yd1}c_yc_{d1} - \rho_{yd1}c_yc_{d1} + \rho_{yd2}c_yc_{d2}\right)\right] > \frac{4C_y^2\alpha - \beta\xi(C_{d1}^2 + C_{d2}^2)}{4[\beta + C_y^2\xi\alpha + \beta\xi(C_{d1}^2 + C_{d2}^2)]}.
\]

From Eqs. (16) and (24)
\[
\text{MSE}(\hat{\mu}_7) > \text{MSE}(\hat{\mu}_{\text{prop}})_{\text{min}} \quad \text{if and only if}
\]
\[
\left( c_y^2 + c_{d1}^2 + c_{d2}^2 + 2 \left( \rho_{yd1} c_y c_{d1} - \rho_{yd1} c_y c_{d1} + \rho_{yd2} c_y c_{d2} \right) \right) > \left( \frac{4C^2_2 \alpha - \beta \xi \left( C_{d1}^2 + C_{d2}^2 \right)^2}{4 \left( \beta + C^2_2 \xi \alpha + \beta \xi \left( C_{d1}^2 + C_{d2}^2 \right) \right)} \right)
\]

From Eqs. (18) and (24)

\[
MSE(\hat{\beta}_s) > MSE(\hat{\beta}_{prop})_{\text{min}} \text{ if and only if }
\]

\[
\bar{y}^2C_y^2[1 - R_{yd1d2}^2] > \left( \frac{4C^2_2 \alpha - \beta \xi \left( C_{d1}^2 + C_{d2}^2 \right)^2}{4 \left( \beta + C^2_2 \xi \alpha + \beta \xi \left( C_{d1}^2 + C_{d2}^2 \right) \right)} \right)
\]

Results and discussion

To examine the dominance and applicability of proposed estimator in simple random sampling, we considered four real data sets from agricultural, fishers and education sectors available in Ahmad et al. [8]. A similar kind of population for reference is given as Appendix A.

Population 1. (Source from education sector)

Let \( y \) represent the number of instructors, \( d_1 \) the total number of primary and secondary school students in Turkey in 2007, which was larger than 11,440.5, and \( d_2 \) the total number of primary and secondary school students in Turkey in 2008, which was greater than 333.1647. The population information about data set is given as:

\[
N = 923, \quad n = 180, \quad \bar{y} = 436.4346, \quad D_1 = 2.6625, \quad D_2 = 3.125, \quad \rho_{yd1} = 0.6904898, \quad \rho_{yd2} = 0.652149, \quad \rho_{d1d2} = 0.8465885, \quad C_{d1} = 1.826732, \quad C_{d2} = 1.641621, \quad C_y = 1.718333
\]

Population 2. (Source from fishers’ sector)

Let \( y \) represent the expected catch by recreational marine fishermen in 1995, \( d_1 \) represent the percentage of fish captured larger than 1000 in 1993, and \( d_2 \) represent the percentage of fish caught greater than 2000 in 1994. The population information about data set is given as:

\[
N = 69, \quad n = 14, \quad \bar{y} = 4514.89, \quad D_1 = 0.7391304, \quad D_2 = 0.5507246, \quad \rho_{yd1} = 0.3966081, \quad \rho_{yd2} = 0.538047, \quad \rho_{d1d2} = 0.6577519, \quad C_{d1} = 0.5984409, \quad C_{d2} = 0.9098277, \quad C_y = 1.350
\]

Population 3. (Source from agricultural sector)

For the 47 districts of Pakistan, let \( y \) represent the tobacco area production in hectares for the year 2009, \( d_1 \) represent the percentage of farms with tobacco cultivation areas greater than 500 ha for the year 2007, and \( d_2 \) represent the percentage of farms with tobacco cultivation areas greater than 800 ha for the year 2008. The population information about data set is given as:

\[
N = 47, \quad n = 10, \quad \bar{y} = 1004.447, \quad D_1 = 0.4255319, \quad D_2 = 0.3829787, \quad \rho_{yd1} = 0.4395989, \quad \rho_{yd2} = 0.4661508, \quad \rho_{d1d2} = 0.9153857, \quad C_{d1} = 1.174456, \quad C_{d2} = 1.283018, \quad C_y = 2.341245
\]

Population 4. (Source from agricultural sector)

Let \( y \) represent the amount of cotton produced in hectares in 2009, \( d_1 \) represent the percentage of farms with cotton cultivation areas greater than 37 ha in 2007, and \( d_2 \) represent the percentage of farms with cotton cultivation areas greater than 35 ha in 2008 for 52 districts in Pakistan. The population information about data set is given as:

\[
N = 52, \quad n = 11, \quad \bar{y} = 50.03846, \quad D_1 = 0.3846154, \quad D_2 = 0.4423077, \quad \rho_{yd1} = 0.7369579, \quad \rho_{yd2} = 0.6935718, \quad \rho_{d1d2} = 0.8877181, \quad C_{d1} = 1.277252, \quad C_{d2} = 1.13384, \quad C_y = 1.421524
\]
Table 1
MSE values estimators.

| Population | Estimator | P₁ | P₂ | P₃ | P₄ |
|------------|----------|----|----|----|----|
|            | ũ₀       | 2515.170 | 2115.180 | 435.363 | 362.664 |
|            | ũ₁       | 1665.169 | 1787.076 | 352.906 | 175.164 |
|            | ũ₂       | 9050.198 | 3274.574 | 736.929 | 1135.734 |
|            | ũ₃       | 1379.541 | 1847.217 | 366.745 | 195.717 |
|            | ũ₄       | 2536.028 | 1295.483 | 279.378 | 338.001 |
|            | ũ₅       | 1315.997 | 1782.466 | 351.230 | 165.699 |
|            | ũ₆       | 5151.870 | 2045.097 | 480.329 | 466.088 |
|            | ũ₇       | 10154.721 | 4937.999 | 870.994 | 1306.264 |
|            | ũ₈       | 1275.439 | 1496.047 | 340.313 | 163.048 |
|            | ũ₀₀ = ũ₈ | 1200.370 | 1290.907 | 206.924 | 106.485 |

Note: Bold number indicate the lowest MSE.

Table 2
Percentage Relative Efficiency (PRE) of estimators.

| Population | Estimator | P₁ | P₂ | P₃ | P₄ |
|------------|----------|----|----|----|----|
|            | ũ₀       | 100 | 100 | 100 | 100 |
|            | ũ₁       | 151.046 | 118.360 | 123.365 | 207.043 |
|            | ũ₂       | 27.791 | 64.594 | 59.078 | 31.932 |
|            | ũ₃       | 182.319 | 114.506 | 118.710 | 185.300 |
|            | ũ₄       | 99.178 | 163.273 | 155.833 | 107.297 |
|            | ũ₅       | 191.123 | 118.666 | 123.954 | 218.870 |
|            | ũ₆       | 48.821 | 103.427 | 90.638 | 77.810 |
|            | ũ₇       | 24.768 | 42.835 | 49.985 | 27.763 |
|            | ũ₈       | 163.048 | 141.384 | 127.929 | 222.428 |
|            | ũ₀₀ = ũ₈  | 209.533 | 163.852 | 210.398 | 340.578 |

Note: Bold number indicate highest PRE.

We compute the MSE values of existing estimates and proposed estimator using Eqs. (2), (4), (6), (8), (10), (12), (14), (16), (18) and (24) and these values are shown in Table 1.

To measure the Percentage Relative Efficiency (PRE), we apply the following formula:

\[
PRE = \frac{MSE(\hat{\mu}_0)}{MSE(t)} \times 100
\]

where \( t = \hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4, \hat{\mu}_5, \hat{\mu}_6, \hat{\mu}_7, \hat{\mu}_8, \) and \( \hat{\mu}_{\text{prop}} \)

The numerical comparison of PRE for existing and proposed estimator is shown in Table 2.

The proposed estimator is compared with the eight estimators as shown in Section 2 and results shown numerically and graphically. It is observed from Table 1, that proposed estimator attains the minimum mean square error 1200.370, 1290.907, 206.924, 106.485 for populations \( P_1, \ldots, P_4 \), respectively. While from Table 2, it is revealed that PRE value of the proposed estimator for all four populations \( P_1, \ldots, P_4 \) is 209.533, 163.852, 210.398 and 340.578, respectively. The PRE of proposed estimator is highest in comparison to the existing estimators as well as all the estimators given in Ahmed et al. [8]. It is observed from Figs. 1 and 2 that proposed estimator is more efficient for population mean estimation in comparison to the existing estimators. In theory of sample surveys, the major objective of any statistician or investigator is to minimize the MSE and maximize the PRE in the estimation to ideally draw an inference about the study population. The proposed estimator is validated with the help of minimum MSE and maximum PRE. It attains the minimum MSE and
Fig. 2. Graphical representation of PRE of estimators w.r.t $\hat{\mu}_0$.

maximum PRE in comparison of the other existing estimators as shown graphically for all four populations.

**Conclusion**

In this article, we proposed an optimum estimator for estimation of population mean in simple random sampling by utilizing the information of the two auxiliary attribute. Mathematical expressions of the proposed estimator i.e. bias, mean squared error (MSE) and minimum mean squared error are derived. Mean square error of proposed estimator is shown in Section 3. The proposed estimator is efficient under the derived conditions discussed in Section 4. The implementation of proposed estimator helps in estimation of précised value of the population mean on the basis of which effective decisions can be made. It is recommended that the proposed estimator may be used in sample surveys in the areas of education, agriculture fisheries and health sciences, etc. Further, the proposed estimator can be extended to utilization of multi auxiliary information under various sampling designs.

**Declaration of Competing Interest**

There is no conflict of interest among the authors.

**Data availability**

Data is available in public domain.

**CRediT authorship contribution statement**

Monika Saini: Conceptualization, Methodology, Validation, Writing – review & editing. Bhatt Ravi Jitendra Kumar: Investigation, Resources, Writing – original draft. Ashish Kumar: Formal analysis, Methodology, Writing – review & editing.

**Appendix A**

Hypothetical Population: (Source: Singh and Choudhary (20), pp. 177)
Let y be the area under wheat crop (in acres) during the year 1974, \(x_1\) be the area under wheat (in acres) 1971 and \(x_2\) be the area under wheat (in acres) in 1973.

### Hypothetical Population: (Source: Singh and Choudhary [19], pp. 177)

| S.N. | Area under wheat (in acres) | 1971 \((x_1)\) | 1973 \((x_2)\) | 1974 \((y)\) |
|------|-----------------------------|----------------|----------------|-----------|
| 1    | 401                         | 70             | 50             |
| 2    | 630                         | 163            | 149            |
| 3    | 1194                        | 320            | 284            |
| 4    | 1170                        | 440            | 381            |
| 5    | 1065                        | 250            | 278            |
| 6    | 827                         | 125            | 111            |
| 7    | 1737                        | 558            | 634            |
| 8    | 1060                        | 254            | 278            |
| 9    | 360                         | 101            | 112            |
| 10   | 946                         | 359            | 355            |
| 11   | 4170                        | 109            | 99             |
| 12   | 1625                        | 481            | 498            |
| 13   | 827                         | 125            | 111            |
| 14   | 96                          | 5              | 6              |
| 15   | 1304                        | 427            | 339            |
| 16   | 377                         | 78             | 80             |
| 17   | 259                         | 75             | 105            |
| 18   | 186                         | 45             | 27             |
| 19   | 1767                        | 564            | 515            |
| 20   | 604                         | 238            | 249            |
| 21   | 700                         | 92             | 85             |
| 22   | 524                         | 247            | 221            |
| 23   | 571                         | 134            | 133            |
| 24   | 962                         | 131            | 144            |
| 25   | 407                         | 129            | 103            |
| 26   | 715                         | 190            | 175            |
| 27   | 845                         | 363            | 335            |
| 28   | 1016                        | 235            | 219            |
| 29   | 184                         | 73             | 62             |
| 30   | 282                         | 62             | 79             |
| 31   | 194                         | 71             | 60             |
| 32   | 439                         | 137            | 100            |
| 33   | 854                         | 196            | 141            |
| 34   | 820                         | 255            | 263            |

The population information about data set is given as:

\[ \delta_1 = 1 \text{ if farms under the wheat crop which have more than 500 acres land during the year 1971, otherwise } \delta_1 = 0. \]

\[ \delta_2 = 1 \text{ if farms under wheat crop which have more than 100 acres land during the year 1973, otherwise } \delta_2 = 0. \]

\[ D_1 \text{ be the proportion of farms under the wheat crop which have more than 500 acres land during the year 1971 and } D_2 \text{ be the proportion of farms under wheat crop which have more than 100 acres land during the year 1973.} \]

Descriptive Statistics:

\[ N = 34, \quad \bar{y} = 199.4, \quad D_1 = 0.6765, \quad D_2 = 0.7353, \quad \rho_{yd1} = 0.599, \quad \rho_{yd2} = 0.599, \]

\[ \rho_{d1d2} = 0.725, \quad S^2_{d1} = 0.225490, \quad S^2_{d2} = 0.200535, \quad S^2_y = 22, 564.6 \]

References

[1] W.G. Cochran, The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce, J. Agric. Sci. 30 (2) (1940) 262–275.
[2] R. Singh, P. Chauhan, N. Sawan, F. Smarandache, Ratio estimators in simple random sampling using information on auxiliary attribute. Auxiliary Information and a priori Values in Construction of Improved Estimators, Infinite Study 1 (2007) 7–17.

[3] S. Malik, R. Singh, An improved estimator using two auxiliary attributes, Appl. Math. Comput. 219 (23) (2013) 10973–10986.

[4] H.P. Singh, S. Kumar, A regression approach to the estimation of the finite population mean in the presence of non-response, Aust. N. Z. J. Stat. 50 (4) (2008) 395–408.

[5] E.J. Ekpenyong, E.I. Enang, Efficient exponential ratio estimator for estimating the population mean in simple random sampling, Hacet. J. Math. Stat. 44 (3) (2015) 689–705.

[6] J. Lu, Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural, biomedical, and power engineering, Math. Probl. Eng. (2017) 1–7, doi:10.1155/2017/8704734.

[7] T. Zaman, C. Kadilar, Novel family of exponential estimators using information of auxiliary attribute, J. Stat. Manag. Syst. 22 (8) (2019) 1499–1509.

[8] S. Ahmad, M. Arslan, A. Khan, J. Shabbir, A generalized exponential-type estimator for population mean using auxiliary attributes, PLoS ONE 16 (5) (2021) e0246947.

[9] J. Mahajan, K. Banal, S. Mahajan, Estimation of crop production using machine learning techniques: a case study of J&K, Int. J. Inf. Technol. 13 (4) (2021) 1441–1448.

[10] A. Kumar, M. Saini, A predictive approach for finite population mean when auxiliary variables are attributes, Thail. Stat. 20 (3) (2022) 575–584.

[11] M.A. Yunusa, P.C. Gidado, Y.M. Ahijjo, A. Danbaba, S.A. Abdulazeer, A. Audu, Modified classes of regression-type estimators of population mean in the presence of auxiliary attribute, Asian Res. J. Math. 18 (1) (2022) 65–89.

[12] V.D. Naik, P.C. Gupta, A note on estimation of mean with known population proportion of an auxiliary character, J. Ind. Soc. Agric. Stat. 48 (2) (1996) 151–158.

[13] S. Kumar, S. Bhogal, Estimation of the population mean in presence of non-response, Commun. Stat. Appl. Methods 18 (4) (2011) 537–548.

[14] H.P. Singh, M.R. Espejo, On linear regression and ratio–product estimation of a finite population mean, J. R. Stat. Soc. Ser. B (12) (2003) 59–67 (The Statistician).

[15] Waheebleb, R., Wheib, K., Andersen, B., & Alshuili, R. (2022). The prospective of Artificial Neural Network (ANN’s) model application to ameliorate management of post disaster engineering projects. Available at SSRN 4180813. doi:10.2139/ssrn.4180813.

[16] A.A. Rajak, Emerging technological methods for effective farming by cloud computing and IoT, Emerg. Sci. J. 6 (5) (2022) 1017–1031.

[17] Z.K. Jabal, T.S. Khayyun, I.A. Alwan, Impact of climate change on crops productivity using MODIS-NDVI time series, Civ. Eng. J. 8 (06) (2012) 1136–1155.

[18] L.H. Wayangkau, Y. Mekiw, R. Rachmat, S. Suwarjono, H. Hariyanto, Utilization of IoT for soil moisture and temperature monitoring system for onion growth, Emerg. Sci. J. 4 (2020) 102–115.

[19] D. Singh, F.S. Chaudhary, Theory and Analysis of Sample Survey Designs, John Wiley & Sons, 1986.

[20] W.G. Cochran, Sampling Techniques, John Wiley & Sons, 2007.

[21] N. Koyuncu, C. Kadilar, Family of estimators of population mean using two auxiliary variables in stratified random sampling, Commun. Stat. Theory Methods 38 (14) (2009) 2398–2417.

[22] J. Shabbir, S. Gupta, Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes, Hacet. J. Math. Stat. 39 (1) (2010) 121–129.

[23] A.M. Abd-Effatah, E.A. El-Sherpieny, S.M. Mohamed, O.F. Abdou, Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute, Appl. Math. Comput. 215 (12) (2010) 4198–4202.

[24] Rather, K.U.I., Jeelani, M.I., Rizvi, S.E.H., & Sharma, M. (2022). A new exponential ratio type estimator for the population means using information on auxiliary attribute. J. Stat. Appl. Pro. Lett. 9, No. 1, 31–42 (2022).

[25] T. Zaman, M. Sagir, M. Şahin, A new exponential estimators for analysis of COVID-19 risk, Concurr. Comput. Pract. Exp. 34 (10) (2022) e6806.