JOURNAL OF STATISTICS, OPTIMIZATION AND DATA SCIENCE
VOL: 1, ISSUE: 2, (DECEMBER 2023), PP: 1-9, ISSN: 2583-9462

POPULATION MEAN ESTIMATION UNDER ADAPTIVE CLUSTER SAMPLING USING RATIO TYPE ESTIMATOR

Arshid Ahmad Bhat[1], Shamshad Ur Rasool[2], Manish Sharma[3], Mohd Younis Shah[4], R K Salgotra[5], Bupesh Kumar[6] ,Mohamad Saleem Mir[7]
[1 3 4]Division of Statistics and Computer Science, Sher-e-Kashmir University of Agricultural Sciences and Technology-Jammu, 180009
[2]Department of Statistics, University Of Kashmir, Srinagar-19006
[4]School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology-Jammu, 180009
[5]Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology-Jammu, 180009
[7]Indian Council of Medical Research, New Delhi
[1]arshidjalalabhat@gmail.com,[2] Srasool92@gmail.com,
[3]manshstat@gmail.com,[4]shahyunu2@gmail.com, [5]rks_2959@rediffmail.com,
[6]bupeshsharma@gmail.com, [7]mir.saleem7@gmail.com

ARTICLE HISTORY
Compiled January 12, 2024

ABSTRACT
In this article, we suggested a generalized class of ratio type estimator to estimate the unknown population mean under Adaptive Cluster Sampling (ACS). The quantitative evaluation is executed to identify the best values of (p and q). The estimator's Mean Squared Error has been estimated up to the first order of approximation and efficiency has been assessed through percentage relative efficiency

KEYWORDS
Adaptive Cluster Sampling; Ratio Type estimator; Rare Population; Auxiliary Variable; Mean Squared Error.

1. Introduction

The adaptive cluster sampling (ACS) approach first suggested by Thompson in 1990 is used to sample rare and hidden clustered populations. Prior to collecting a sample, ACS requires the neighbourhood of an initially selected unit as well as the condition of interest 'C' for modifying the nearby units. Both of these requirements are open for interpretation to the researcher. The condition of interest 'C' is generally written as $y_i > 0$, where y_i signifies the i^{th} unit or observation on the variable of interest 'Y' and the neighbourhood typically contains the four basic directional units of a selected unit. A population of size N is sampled initially using probability sampling approach yielding a sample size of n. In this article simple random sampling without replacement (SRSWOR) is used to obtain the initial sample of size n. The adaptive sample has been generated based on the pre specified condition 'C'. Every unit around the initially sampled unit that satisfies condition of interest 'C' is selected from its first order (four unit) neighbourhood. The first order neighbourhood of any unit that satisfies the interest $y_i > 0$ requirements is taken around the initial unit until there are no longer any units that meet the

CONTACT Arshid Ahmad Bhat[a]. Email: arshidjalalbhat@gmail.com

interest requirement $y_i > 0$. There are some observations in the neighbourhood that match desired condition C and some observations that don't. Networks are defined as observations that satisfy criterion C, whereas edge units are defined as observations that do not. A cluster is created by these networks and edge units. The entire population can be divided into exhaustive sets of networks. Due to overlapping edge units the clusters are not always disjoint but the networks are disjoint as a result the final sample is made up of the original sample and the adaptive samples. It is widely known that the variance of the estimator of the population's parameter(s) of interest that are associated (either positively or negatively) with the variable of interest can be greatly reduced by the use of an auxiliary variable. The traditional ratio and product type estimators proposed by Cochran (1940) and Robson (1957) respectively are the fundamental estimators for the estimation of finite population mean in simple random sampling. Subsequent researchers have suggested the modified ratio and product type estimators including Sisodia and Dwivedi (1981), Bahl and Tuteji (1991), Upadhyay and Singh (1999), Singh and Tailor (2003), Kadilar and Cingi (2003), Sharma and Bhatnagar (2008), Yan and Tian (2010), Chutiman (2013), Yadav and Kadilar (2013), Shahzad and Hanif (2016), Jeelani et al. (2017), Kumar et al. (2018), Hussain et al. (2021), and Arshid et al. (2022).

2. Methodology

The selection of a random sample is the first stage in the ACS process. This sample is chosen randomly using conventional sampling methods, ideally by simple random sampling without replacement (SRSWOR). In ACS, the surrounding units of the initially selected sample are included after satisfying a predefined condition 'C'. The neighbourhood which is defined as the four traditional directional units (East, West, North and South) is checked to see if the units satisfy the specified condition or not, if it does it is included to the sample otherwise not. Any extra unit in the neighbourhood that is adaptively after satisfying the condition is likewise included in the sample. The units that don't fulfil the criteria are positioned in close proximity to the units that were adaptively added. Every unit that was initially selected and its neighbour that satisfies the condition make up a network. The units on the edge are those that don't fit the requirements. The network and associated edge units are referred to as a cluster. If the units selected in the initial sample does not satisfy the predetermined condition then there is only one unit in the network. The original sample plus all of the units that were adaptively added together make up the network's final sample size. 1 show a cluster in which the unit with a star (*) is the first selected unit. The units around the starred unit shaded with blue form a network(ψ_i). The bold units (zeroes) are edge units. Together the edge units and the network form a cluster. This strategy provides a better estimate of the population parameter when the population is unusual, concentrated, and challenging to reach.

Let the usual finite population consists N distinct units labelled from $1, 2, \ldots, N$. the variables$y_i$and $X_i (i = 1, 2, 3. \ldots., N)$ denote the i^{th} value for the survey and auxiliary variables respectively, 'n' denote the initial sample size. Let the population is divided into K exhaustive networks where (ψ_i) denotes the network that includes i units with $m_i$number of units in the i^{th}network. The mean, standard deviation, coefficient of variation, correlation coefficient and covariance of survey and auxiliary variable at network level is denoted by

$$\overline{w}_y, \overline{w}_x, \sigma_{wy}, \sigma_{wx}, C_{wy}, C_{wx}, \rho_{wxwy}, \sigma_{wywx}$$

The parameters of study and auxiliary variables are as follows:

Study Variable

Mean: $\mu_y = \frac{1}{N} \sum_{i=1}^{N} y_i$

Variance: $\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \mu_y)^2$

Auxiliary Variable

Mean: $\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i$

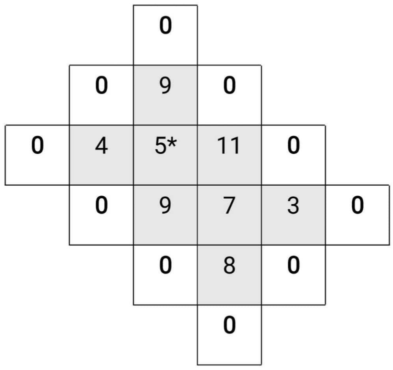

Figure 1. The shaded units 4, 9, 5, 911, 17, 8, 3 form the network, the bold units like (0s) are the edge units, and together these units form a cluster.

Variance: $\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2$
Let $w_{yi} = \frac{1}{m_i}\sum_{j\in\phi_i} y_i$ and $w_{xi} = \frac{1}{m_i}\sum_{j\in\phi_i} x_i$ be the transformed survey and auxiliary variables in the ith network, respectively.
The transformed population parameters of study and auxiliary variables are:

Study variable

$$\text{Mean: } \mu_{\boldsymbol{wy}} = \mu_y = N^{-1}\sum_{i=1}^{N} w_{yi}$$

$$\text{Variance: } \sigma^2_{\boldsymbol{wy}} = (N-1)^{-1}\sum_{i=1}^{N}(w_{yi} - \mu_{wy})^2$$

Auxiliary variable

$$\text{Mean: } \mu_{\boldsymbol{wx}} = \mu_x = N^{-1}\sum_{i=1}^{N} w_{xi}$$

$$\text{Variance: } \sigma^2_{\boldsymbol{wx}} = (N-1)^{-1}\sum_{i=1}^{N}(w_{xi} - \mu_{wx})^2$$

Study variable

$$\text{Mean: } \bar{w}_y = n^{-1}\sum_{i=1}^{n} w_{yi}$$

$$\text{Variance: } s^2_{wy} = (n-1)^{-1}\sum_{i=1}^{n}(w_{yi} - \bar{w}_y)^2$$

Auxiliary variable

$$\text{Mean: } \bar{w}_x = n^{-1}\sum_{i=1}^{n} w_{xi}$$

$$\text{Variance: } s^2_{wx} = (n-1)^{-1}\sum_{i=1}^{n}(w_{xi} - \bar{w}_x)^2$$

3. Results and discussion:

The ratio estimator is a statistical parameter defined as the ratio of means of two random variables. Several authors have used the ratio estimator in order to increase the precision of the estimators. The pioneer work has been done by Cochran (1940), later on Sisodia and Dwivedi (1981), Upadhyay and Singh (1999), Singh and Tailor (2003), Kadilar and Cingi (2003),Yan and Tian (2010) proposed modified ratio type estimators by using coefficient of variation (CV), kurtosis, correlation coefficient, skewness etc. as the auxiliary variables. Some of the existing estimators and their expressions of mean square error derived upto first order of approximation are enlisted in the following table;

Table 1. Existing estimators and their MSE

Researcher	**Estimator**	**Mean Square Error**
Cochran (1940)	$Y\hat{\bar{R}} = t_1 = \bar{y}\left(\frac{\mu_x}{\bar{x}}\right)$	$\theta\bar{Y}^2(C_y^2 + C_x^2 - 2\rho_{xy}C_xC_y)$
Sisodia and Dwivedi (1981)	$t_2 = \bar{y}\left(\frac{\mu_x+C_x}{\bar{x}+C_x}\right)$ Where $\theta_1 = \frac{\mu_x}{\mu_x+C_x}$	$\theta\mu_y^2(C_y^2 + \theta_1 C_x^2 - 2\theta_1\rho_{xy}C_xC_y)$
Upadhyay and Singh (1999)	$t_3 = \bar{y}\left(\frac{\beta_{2(x)}\mu_x+C_x}{\beta_{2(x)}\bar{x}+C_x}\right)$ $t_4 = \bar{y}\left(\frac{C_x\mu_x+\beta_{2(x)}}{C_x\bar{x}+\beta_{2(x)}}\right)$ Where $\theta_2 = \frac{\beta_{2(x)}\mu_x}{\beta_{2(x)}\mu_x+C_x}$	$\theta\mu_y^2(C_y^2 + \theta_2^2 C_x^2 - 2\theta_2\rho_{xy}C_xC_y)$ $\theta\mu_y^2(C_y^2 + \theta_3^2 C_x^2 - 2\theta_3\rho_{xy}C_xC_y)$ Where $\theta_3 = \frac{C_x\mu_x}{C_x\mu_x+\beta_{2(x)}}$
Singh and Tailor (2003)	$t_5 = \bar{y}\left(\frac{\mu_x+\rho_{xy}}{\bar{x}+\rho_{xy}}\right)$ Where $\theta_4 = \frac{\mu_x}{\mu_x+\rho_{xy}}$	$\theta\mu_y^2(C_y^2 + \theta_4^2 C_x^2 - 2\theta_4\rho_{xy}C_xC_y)$
Kadilar and Cingi (2003)	$t_6 = \bar{y}\left(\frac{\mu_x^2}{\bar{x}^2}\right)$	$\theta\mu_y^2(C_y^2 + 4C_x^2 - 4\rho_{xy}C_xC_y)$
Yan and Tian (2010)	$t_7 = \bar{y}\left(\frac{\beta_{2(x)}\mu_x+\beta_{1(x)}}{\beta_{2(x)}\bar{x}+\beta_{1(x)}}\right)$ $t_8 = \bar{y}\left(\frac{\beta_{1(x)}\mu_x+\beta_{2(x)}}{\beta_{1(x)}\bar{x}+\beta_{2(x)}}\right)$ Where $\theta_5 = \frac{\beta_{2(x)}\mu_x}{\beta_{2(x)}\mu_x+\beta_{1(x)}}$	$\theta\mu_y^2(C_y^2 + \theta_5^2 C_x^2 - 2\theta_5\rho_{xy}C_xC_y)$ $\theta\mu_y^2(C_y^2 + \theta_6^2 C_x^2 - 2\theta_6\rho_{xy}C_xC_y)$ Where $\theta_6 = \frac{\beta_{1(x)}\mu_x}{\beta_{1(x)}\mu_x+\beta_{2(x)}}$

The usual notations used in the above derivations are defined as follows;

$$\theta = \frac{N-n}{Nn} = \frac{1-f}{n}, \quad f = \frac{n}{N}$$
$$C_y = \frac{S_y}{\mu_y}, \quad C_x = \frac{S_x}{\mu_x}$$
$$s_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu_y)^2, \quad s_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$$
$$\rho_{xy} = \frac{S_{xy}}{\sqrt{S_xS_y}}, \quad S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$$

But, there are situations when the population is rare and a researcher can't get information about the population as a whole, but they can get information about the clusters. In handling such situations ACS is used. Thompson (1990) proposed the unbiased mean estimator based on the modification of Hansen-Hurwitz (HH) estimator as;

$$\hat{\mu}_y^R = t_9 = \frac{1}{n}\sum_{i=1}^{n} w_{yi}$$

The variance of $\hat{\mu}_y^R = t_9$ is given by:

$$\hat{\mu}_y^R = \theta\mu_{wy}^2 C_{wy}^2 \quad (1) \tag{1}$$

Where $\theta = \frac{N-n}{Nn} = \frac{1-f}{n}$ and $C_{wy}^2 = \frac{S_{wy}^2}{\mu_{wy}^2}$. Subsequently, Dryver and Chao (2007) proposed the biased ratio type estimators on the modification of HH type estimator for the estimation of

population mean under ACS as;
$t_{10} = \frac{\bar{w}_y}{\bar{w}_x}\mu_x$.

The mean square error (MSE) of t_{10} up to $O(n^{-1})$ is given by:

$$\text{MSE } t_{10} \approx \theta\mu_{wy}^2(C_{wy}^2 + C_{wx}^2 - 2\rho_{wxwy}C_{wx}C_{wy}) \tag{2}$$

Later Chutiman (2013) proposed ratio-type estimators on the basis of Dryver and Chao (2007) as:

$$t_{11} = \frac{\bar{w}_y(\mu_x + C_{wx})}{\bar{w}_x + C_{wx}}$$
$$t_{12} = \frac{\bar{w}_y(\mu_{x\beta_{2(wx)}} + C_{wx})}{\bar{w}_{x\beta_{2(wx)}} + C_{wx}}$$
$$t_{13} = \frac{\bar{w}_y(\mu_x + \beta_{2(wx)})}{\bar{w}_x + \beta_{2(wx)}}$$

The mean square error (MSE) of t_{11} up to $O(n^{-1})$ is given by:

$$\text{MSE}(t_{11}) \approx \theta\mu_{wy}^2(C_{wy}^2 + \theta_{w7}^2C_{wx}^2 - 2\theta_{w7}\rho_{wxwy}C_{wx}C_{wy}) \tag{3}$$

Where $\theta_{w7} = \frac{\mu_{wx}}{\mu_{wx}+C_{wx}}$.

The MSE of t_{12} up to $O(n^{-1})$ is given by:

$$\text{MSE}(t_{12}) \approx \theta\mu_{wy}^2(C_{wy}^2 + \theta_{w8}^2C_{wx}^2 - 2\theta_{w8}\rho_{wxwy}C_{wx}C_{wy}) \tag{4}$$

Where $\theta_{w8} = \frac{\mu_x\beta_{2(wx)}}{\mu_x\beta_{2(wx)}+C_{wx}}$.

The MSE of t_{13} up to $O(n^{-1})$ is given by:

$$\text{MSE}(t_{13}) \approx \theta\mu_{wy}^2(C_{wy}^2 + \theta_{w9}^2C_{wx}^2 - 2\theta_{w9}\rho_{wxwy}C_{wx}C_{wy}) \tag{5}$$

Where $\theta_{w9} = \frac{\mu_x}{\mu_x+\beta_{2(wx)}}$.
Subhash et al. (2016) proposed ratio type estimators on the basis of Dryver and Chao (2007) as: The ratio-type estimator t_{14} is given by:

$$t_{14} = \bar{w}_y\left(\frac{\mu_x^2}{\bar{w}_x}\right)$$

The ratio-type estimator t_{15} is given by:

$$t_{15} = \bar{w}_y\left(\frac{\beta_{1(wx)}\mu_x + \beta_{2(wx)}}{\beta_{1(wx)}\bar{w}_x + \beta_{2(wx)}}\right)$$

The mean square error (MSE) of t_{14} up to $O(n^{-1})$ is given as:

$$\text{MSE}(t_{14}) \approx \theta\mu_y^2(C_{wy}^2 + 4C_{wx}^2 - 4\rho_{wxwy}C_{wx}C_{wy}) \tag{6}$$

The MSE of t_{15} up to $O(n^{-1})$ is given as:

$$\text{MSE}(t_{15}) \approx \theta\mu_y^2(C_{wy}^2 + \theta_{w7}^2C_{wx}^2 - 2\theta_{w7}\rho_{wxwy}C_{wx}C_{wy}) \tag{7}$$

Where $\theta_{w10} = \frac{\mu_x\beta_{1(wx)}}{\mu_x\beta_{1(wx)}+\beta_{2wx}}$.

Considering the generalized class of ratio-type estimator under ACS as:

$$\mu_y^r = t'_{(p,q)} = \bar{w}_y \left(\frac{\mu_x + p\rho_{wxwy}}{\bar{w}_x + q\rho_{wxwy}} \right)$$

The MSE of $\mu_y^r = t'_{(p,q)}$ is given as:

$$\text{MSE}(\mu_y^r) = t'_{(p,q)} = \mu_y^2 \left[(\theta_{w11} - 1)^2 + \theta_{w11}^2 \theta C_{wy}^2 + \theta_{w11}\theta_{w12}^2 \theta C_{wx}^2 (3\theta_{w11} - 2) + \theta_{w11}\theta_{w12}\theta\rho_{wxwy} C_{wx} C_{wy} (2 - 4\theta_{w11}) \right] \quad (8)$$

Where $\theta_{w11} = \frac{\mu_x + p\rho_{wxwy}}{\mu_x + q\rho_{wxwy}}$ and $\theta_{w12} = \frac{\mu_x}{\mu_x + q\rho_{wxwy}}$.

Special cases:

- Substituting $p = 1$, $q = 7$ in the proposed estimator $\mu_y^r = t'_{(p,q)}$, we get the estimator $\hat{\mu}_y^r = t'_{1,7} = \bar{w}_y \left(\frac{\mu_x + C_x}{\bar{x} + C_x} \right)$, which is Sisodia and Dwivedi (1981) estimator and is a special class of the proposed estimator.
- Substituting $p = 2$, $q = 18$, in the proposed estimator $\mu_y^r = t'_{(p,q)}$, we get the estimator $\hat{\mu}_y^r = t'_{2,18} = \frac{C_x \mu_x + \beta_{2(x)}}{C_x \bar{x} + \beta_{2(x)}}$, which is Upadhyay and Singh (1999) and is a special class of the proposed estimator.

4. Theoretical Comparison:

The algebraic expressions showed that the MSE of the proposed estimator is less than the literature-based estimators. The proposed estimator is better than Sisodia and Dwivedi (1981) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_2) \\ &\left[(\theta_{w11} - 1)^2 + \theta_{w11}^2 \theta C_{wy}^2 + \theta_{w11}\theta_{w12}^2 \theta C_{wx}^2 (3\theta_{w11} - 2) + \theta_{w11}\theta_{w12}\theta\rho_{wxwy} C_{wx} C_{wy} (2 - 4\theta_{w11}) \right] \\ &\quad - \theta(C_y^2 + \theta_1 C_x^2 - 2\theta_1 \rho_{xy} C_x C_y) < 0 \end{aligned} \quad (9)$$

The proposed estimator is better than Upadhyay and Singh (I) (1999) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_3) \\ &\left[(\theta_{w11} - 1)^2 + \theta_{w11}^2 \theta C_{wy}^2 + \theta_{w11}\theta_{w12}^2 \theta C_{wx}^2 (3\theta_{w11} - 2) + \theta_{w11}\theta_{w12}\theta\rho_{wxwy} C_{wx} C_{wy} (2 - 4\theta_{w11}) \right] \\ &\quad - \theta(C_y^2 + \theta_2^2 C_x^2 - 2\theta_2 \rho_{xy} C_x C_y) < 0 \end{aligned} \quad (10)$$

The proposed estimator is better than Upadhyay and Singh (II) (1999) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_4) \\ &\left[(\theta_{w11} - 1)^2 + \theta_{w11}^2 \theta C_{wy}^2 + \theta_{w11}\theta_{w12}^2 \theta C_{wx}^2 (3\theta_{w11} - 2) + \theta_{w11}\theta_{w12}\theta\rho_{wxwy} C_{wx} C_{wy} (2 - 4\theta_{w11}) \right] \\ &\quad - \theta(C_y^2 + \theta_3^2 C_x^2 - 2\theta_3 \rho_{xy} C_x C_y) < 0 \end{aligned} \quad (11)$$

The proposed estimator is better than Singh and Tailor (2003) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_5) \\ &\left[(\theta_{w11} - 1)^2 + \theta_{w11}^2 \theta C_{wy}^2 + \theta_{w11}\theta_{w12}^2 \theta C_{wx}^2 (3\theta_{w11} - 2) + \theta_{w11}\theta_{w12}\theta\rho_{wxwy} C_{wx} C_{wy} (2 - 4\theta_{w11}) \right] \\ &\quad - \theta(C_y^2 + \theta_4^2 C_x^2 - 2\theta_4 \rho_{xy} C_x C_y) < 0 \end{aligned} \quad (12)$$

The proposed estimator is better than Kadilar and Singi (2003) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_6) \\ &\big[(\theta_{w11}-1)^2+\theta_{w11}^2\theta C_{wy}^2+\theta_{w11}\theta_{w12}^2\theta C_{wx}^2+\theta_{w11}\theta_{w12}\theta\rho_{wxwy}C_{wx}C_{wy}(2-4\theta_{w11})\big] \\ &\quad -\theta(C_y^2+4C_x^2-4\rho_{xy}C_xC_y)<0 \end{aligned} \tag{13}$$

The proposed estimator is better than Yan and Tian (2010) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_8) \\ &\big[(\theta_{w11}-1)^2+\theta_{w11}^2\theta C_{wy}^2+\theta_{w11}\theta_{w12}^2\theta C_{wx}^2(3\theta_{w11}-2)+\theta_{w11}\theta_{w12}\theta\rho_{wxwy}C_{wx}C_{wy}(2-4\theta_{w11})\big] \\ &\quad -\theta(C_y^2+\theta_6^2C_x^2-2\theta_6\rho_{xy}C_xC_y)<0 \end{aligned} \tag{14}$$

The proposed estimator is better than Chutiman (I) (2013) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_{11}) \\ &\big[(\theta_{w11}-1)^2+\theta_{w11}^2\theta C_{wy}^2+\theta_{w11}\theta_{w12}^2\theta C_{wx}^2(3\theta_{w11}-2)+\theta_{w11}\theta_{w12}\theta\rho_{wxwy}C_{wx}C_{wy}(2-4\theta_{w11})\big] \\ &\quad -\theta(C_{wy}^2+\theta_7^2C_{wx}^2-2\theta_7\rho_{wxwy}C_{wx}C_{wy})<0 \end{aligned} \tag{15}$$

The proposed estimator is better than Chutiman (III) (2013) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_{13}) \\ &\big[(\theta_{w11}-1)^2+\theta_{w11}^2\theta C_{wy}^2+\theta_{w11}\theta_{w12}^2\theta C_{wx}^2(3\theta_{w11}-2)+\theta_{w11}\theta_{w12}\theta\rho_{wxwy}C_{wx}C_{wy}(2-4\theta_{w11})\big] \\ &\quad -\theta(C_{wy}^2+\theta_9^2C_{wx}^2-2\theta_9\rho_{wxwy}C_{wx}C_{wy})<0 \end{aligned} \tag{16}$$

The proposed estimator is better than Subhash et al. (I) (2016) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_{14}) \\ &\big[(\theta_{w10}-1)^2+\theta_{w10}^2\theta C_{wy}^2+\theta_{w10}\theta_{w11}^2\theta C_{wx}^2(3\theta_{w10}-2)+\theta_{w10}\theta_{w11}\theta\rho_{wxwy}C_{wx}C_{wy}(2-4\theta_{w10})\big] \\ &\quad -\theta(C_{wy}^2+4C_{wx}^2-4\rho_{wxwy}C_{wx}C_{wy})<0 \end{aligned} \tag{17}$$

The proposed estimator is better than Subhash et al. (II) (2016) if:

$$\begin{aligned} &\text{MSE}(t'_{(p,q)}) < \text{MSE}(t_{15}) \\ &\big[(\theta_{w11}-1)^2+\theta_{w11}^2\theta C_{wy}^2+\theta_{w11}\theta_{w12}^2\theta C_{wx}^2+\theta_{w11}\theta_{w12}\theta\rho_{wxwy}C_{wx}C_{wy}(2-4\theta_{w11})\big] \\ &\quad -\theta(C_{wy}^2+\theta_{w10}^2C_{wx}^2-2\theta_{w10}\rho_{wxwy}C_{wx}C_{wy})<0 \end{aligned} \tag{18}$$

5. Numerical Illustration

In order to evaluate the performance of proposed estimator, the simulated data of x-values and y-values from Chutiman and Kumphon (2008) were studied. The data statistics is given in table 2, the MSE values of proposed estimators for different values of p and q are given in Table 3, and percentage relative efficiency is given in table 4. From the above table it is found that the proposed estimator is most efficient than all existing estimators.

Table 2. Summary Statistics

μ_y	1.2225	μ_x	0.550
N	400	n	20
$\beta_1(wx)$	7.953	$\beta_2(wx)$	91.369
C_{wx}	3.510	C_{wy}	2.914
S_{wx}	1.948	S_{wy}	3.562
S_{wxwy}	6.428	ρ_{wxwy}	0.926
ρ_{xy}	0.910	C_x	4.325
C_y	4.131	θ_1	0.114
θ_2	0.876	θ_3	0.042
θ_4	0.379	θ_6	0.817
θ_7	0.064	β_{2x}	55.090
θ_{w1}	0.137	θ_{w2}	0.935
θ_{w3}	0.006	θ_{w4}	0.375
θ_{w6}	0.864	θ_{w7}	0.046
θ_4	0.379	β_{1x}	6.832
$\beta_{1(wx)}$	7.953	$\beta_{2(wx)}$	91.369

Table 3. MSE Values of Proposed Estimators for Different Values of p and q

$t'_{(p,q)}$	$\boldsymbol{\theta_{w16}}$	$\boldsymbol{\theta_{w17}}$	**MSE**
$t'_{(2,2)}$	1.000	0.230	0.339
$t'_{(2,3)}$	0.722	0.166	0.361
$t'_{(3,4)}$	0.782	0.130	0.366
$t'_{(4,5)}$	0.821	0.107	0.382
$t'_{(5,6)}$	0.848	0.090	0.399

Table 4. Percentage Relative Efficiency (PRE) of the Proposed Estimator with Respect to Existing Estimators (EE),where the MSE of proposed estimator (PE) = 0.339, least among the other proposed estimators

Existing Estimator (EE)	**MSE of EE**	**PRE (%)**
$t_2 = \bar{y}\left(\frac{\mu_x+C_x}{\bar{x}+C_x}\right)$	0.966	242.105
$t_3 = \bar{y}\left(\frac{\beta_{2(x)}\mu_x+C_x}{\beta_{2(x)}\bar{x}+C_x}\right)$	0.407	120.005
$t_4 = \bar{y}\left(\frac{C_x\mu_x+\beta_{2(x)}}{C_x\bar{x}+\beta_{2(x)}}\right)$	1.117	279.949
$t_5 = \bar{y}\left(\frac{\mu_x+\rho_{xy}}{\bar{x}+\rho_{xy}}\right)$	0.527	132.080
$t_6 = \bar{y}\left(\frac{\mu_x^2}{\bar{x}^2}\right)$	1.904	477.192
$t_8 = \bar{y}\left(\frac{\beta_{1(x)}\mu_x+\beta_{2(x)}}{\beta_{1(x)}\bar{x}+\beta_{2(x)}}\right)$	1.068	267.699
$t_{11} = \bar{w}_y\left(\frac{\mu_x+C_{wx}}{\bar{w}_x+C_{wx}}\right)$	0.432	108.270
$t_{13} = \bar{w}_y\left(\frac{\mu_x+\beta_{2(wx)}}{\bar{w}_x+\beta_{2(wx)}}\right)$	0.595	149.122
$t_{14} = \bar{w}_y\left(\frac{\mu_x^2}{\bar{w}_x}\right)$	1.010	253.132
$t_{15} = \bar{w}_y\left(\frac{\beta_{1(wx)}\mu_x+\beta_{2(wx)}}{\beta_{1(wx)}\bar{w}_x+\beta_{2(wx)}}\right)$	0.542	135.839

6. conclusion

It is concluded that the generalized class of proposed estimator at different values of p and q performs better than the existing estimators compared on the basis of percentage relative efficiency. Among the different cases developed on the base of proposed estimator the proposed estimator is best and should be preferred for the estimation of population mean when population is rare and clustered. The expression for MSE of the proposed estimator has been derived. It has been observed that the proposed is more efficient theoretically and empirically than Sisodia and Dwivedi (1981), Upadhyay and Singh (1999), Singh and Tailor (2003), Kadilar and Singi (2003), Yan and Tian (2010), Chutiman (2013), Subhash et al. (2016)

References

[1] Chutiman, N. “Adaptive Cluster Sampling using Auxiliary Variable”. J. Math. Stat., 9, pp. 249-255 (2013).

[2] Cochran, W.G. “The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce”. Journal Agricultural Science, 30, pp. 262-275 (1940).

[3] Dryver, A. L. & Chao, C.T. “Ratio estimators in adaptive cluster sampling”. Environmetrics, 18, pp. 607–620 (2007).

[4] Hansen, M. M. & Hurwitz, W.N. “On the theory of sampling from a finite population”. Ann Math Stat, 14, pp. 333–362 (1943).

[5] Horvitz, D.G. & Thompson, D. J. “A generalization of sampling without replacement from a finite universe”. Journal of American Statistical Association, 47, pp. 663–685 (1952).

[6] Hussain, S., Sharma, M. and Chandra, H. “Optimum exponential ratio type estimator for estimating the population mean”. J. Stat. Appl. Prob. Lett. 8(2), pp. 73-82 (2021).

[7] Jeelani, M. I., Rizvi, S.E.H., Sharma, M. K., Mir, S. A., Raja, S.A., Maqbool, S. A., Nazir, N., & Jeelani, F. “Modified Ratio Estimator under Rank Set Sampling”. Revista Investigacion Operacional, 38(1), pp.103-106 (2017).

[8] Kadilar, B. & Cingi, H. “A study on the chain ratio-type estimator. Hacettepe Journal of Mathematics and Statistics”, pp. 32, 105-108 (2003).

[9] Kumar, B., Sharma, M. K., Rizvi, S.E.H. and Jeelani, M. I. “Estimation of Population Mean through Improved Class of Ratio Type Estimators Using Auxiliary Information”. Journal of Reliability and Statistical Studies, 11(1), pp. 01-07 (2018).

[10] Robson, D.S. “Applications of multivariate polykays to the theory of unbiased ratio-type estimation”. Journal American Statistical Association, 52, pp. 511-522 (1957).

[11] Sharma, M. & Bhatnagar, S. “A General Class Of Estimator For The Estimation Of Population Mean Using Auxiliary Information”. Journal of the Indian Statistical Association, 46(2), pp. 155-161 (2008).

[12] Sisodia, V. S. & Dwivedi, V.K. “A modified ratio estimator using coefficient of variation of auxiliary variable”. J. Indian Soc. Agric. Statist., 33, pp. 13-18 (1981).

[13] Thompson, S.K. “Adaptive cluster sampling”. Journal of American Statistical Association, 85, 1050–1059 (1990).

[14] Upadhyaya, N & Singh, H.P. “Use of transformed auxiliary variable in estimating the finite population mean”. Biometri. J., 41, 627-636 (1999).

[15] Yan & Tian, B. “Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable”. ICICA 2010, Part II, CCIS 106, pp. 103-110 (2010).