

Generalized Ratio Type Estimator under Adaptive Cluster Sampling

Arshid Ahmad Bhat*¹ Manish Sharma² Mohd Younis Shah³ Mehraj Ud Din Bhat⁴

¹Division of Statistics and Computer Science, SKUAST- Jammu, arshidjalalbhat@gmail.com

²Division of Statistics and Computer Science, SKUAST- Jammu, manshstat@gmail.com

³Division of Statistics and Computer Science, SKUAST- Jammu, syeedunis121@gmail.com

⁴Jammu and Kashmir Entrepreneurship Development Institute of Pampore, J&K, mehrajbhat6414@gmail.com

Abstract: The Adaptive Cluster Sample (ACS) sampling design is employed for the evaluation of scarce, concentrated, vulnerable, and complicated population parameters. Generalized ratio type estimator under ACS has been proposed in this paper. The theoretical formulas of the suggested estimator's mean square error is generated up to first order. Using the simulated x- and y-values from Chutiman and Kumphon (2008), the proposed estimator's effectiveness is empirically evaluated with certain known ratio and ACS estimators. Furthermore, the empirical optimum values of p and q are determined. The results indicated that when there is a positive correlation between the study variable and auxiliary variable then the suggested estimate under ACS is a reliable alternative to other estimators.

Index Terms: ACS; Ratio Estimator; Bias; Mean Square Error; Efficiency.

I. INTRODUCTION

It is extremely complicated to get the population mean of uncommon animals, plants, hidden, or indigenous species, uneven mineral mining, pollution concentrations, noise disturbances, drug users, HIV infected patients, thieves, and hot spot inquiries. In these cases Adaptive Cluster Sampling is the effective option of sampling design with the use of ratio type estimators. Thompson suggested this sampling procedure in 1990. In cases, when the area is broad, heterogeneous, and responsive to change, the design is most robust. The estimators in ACS are classified into two categories: the first category includes modified Hansen and Hurwitz (1943) and Horvitz and Thompson (1952) estimators while the second category includes ratio-type estimators that take use of auxiliary data. In ratio method of estimation the auxiliary variable plays an important role to get the precise estimate of the study variable where the auxiliary variable is correlated with the study variable positively and satisfying the relation $\frac{1C_x}{2C_y} \leq \rho \leq 1$. Several authors like Sharma and Bhatnagar (2008), Jeelani et al., (2017), Kumar et

al., (2018) Hussain et al., (2021) have presented estimators for the estimation of population mean that yields better results than the usual sample mean. In ACS, The initial sample in ACS is collected using a probability sampling design.. A specific condition is defined in advance in order to adopt the neighbouring units. If the specified condition is satisfied in the initially selected samples, the neighbouring units are examined and added to the sample, this process continues until the neighbouring units don't satisfy the condition. The units that meet the condition form the network and the units that don't meet the condition are called edge units. The collection of network and edge units is called as cluster. The neighbourhood are the spatially adjacent units in the east, west, north and south of the selected units that satisfies the pre-specified condition. The figure 1 illustrates an example of a cluster having network with its associated edge units as four spatially adjacent units. The unit with a star (*) is an initially selected unit, the condition of adaption is pre-defined as a unit greater than or equal to 1 will be added to the network. The unit which are in the east, west, north and south of the initially selected sample are known as first order neighbourhood. The units in the shaded area form a network while the units in bold numbers are respective edge units of a network. The network with its associated edge units makes up a cluster.

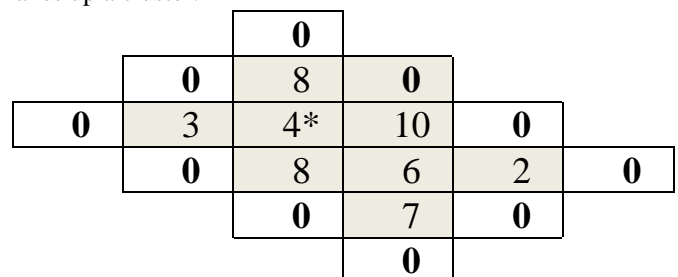


Fig. 1. Cluster (Network and Edge Units)

II. MANUSCRIPT ORGANIZATION

The ratio estimator is defined as the division 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). The proposed estimator and the mean square error are given as $\widehat{Y}_R = \bar{y} \left(\frac{\mu_x}{\bar{x}} \right)$ and $MSE(\widehat{Y}_R) = \theta \bar{y}^2 (C_y^2 + C_x^2 - 2\rho_{xy} C_x C_y)$. With due course of time Sisodia and Dwivedi (1981) used coefficient of variation as an auxiliary variable and proposed ratio estimator as:

$$t_1 = \bar{y} \left(\frac{\mu_x + C_x}{\bar{x} + C_x} \right)$$

$$MSE(t_1) = \theta \mu_y^2 (C_y^2 + \theta_1 C_x^2 - 2\theta_1 \rho_{xy} C_x C_y) \quad (1)$$

where $\theta_1 = \frac{\mu_x}{\mu_x + C_x}$

Upadhyay and Singh (1999) made the use of auxiliary information as kurtosis and coefficient of variation and proposed the estimator as:

$$t_2 = \bar{y} \left(\frac{\beta_{2(x)} \mu_x + C_x}{\beta_{2(x)} \bar{x} + C_x} \right)$$

$$MSE(t_2) = \theta \mu_y^2 (C_y^2 + \theta_2^2 C_x^2 - 2\theta_2 \rho_{xy} C_x C_y) \quad (2)$$

where $\theta_2 = \frac{\beta_{2(x)} \mu_x}{\beta_{2(x)} \mu_x + C_x}$

$$t_3 = \bar{y} \left(\frac{C_x \mu_x + \beta_{2(x)}}{C_x \bar{x} + \beta_{2(x)}} \right)$$

$$MSE(t_3) = \theta \mu_y^2 (C_y^2 + \theta_3^2 C_x^2 - 2\theta_3 \rho_{xy} C_x C_y) \quad (3)$$

where $\theta_3 = \frac{C_x \mu_x}{C_x \mu_x + \beta_{2(x)}}$

Singh and Tailor (2003) suggested ratio type estimator with the use of correlation coefficient as:

$$t_4 = \bar{y} \left(\frac{\mu_x + \rho_{xy}}{\bar{x} + \rho_{xy}} \right)$$

$$MSE(t_4) = \theta \mu_y^2 (C_y^2 + \theta_4^2 C_x^2 - 2\theta_4 \rho_{xy} C_x C_y) \quad (4)$$

where $\theta_4 = \frac{\mu_x}{\mu_x + \rho_{xy}}$

Kadilar and Cingi (2003) suggested an estimator as:

$$t_5 = \bar{y} \frac{\mu_x^2}{\bar{x}^2}$$

$$MSE(t_5) = \theta \mu_y^2 (C_y^2 + 4C_x^2 - 4\rho_{xy} C_x C_y) \quad (5)$$

Yan and Tian (2010) proposed ratio type estimator by taking skewness and kurtosis into account as auxiliary information:

$$t_6 = \bar{y} \left(\frac{\beta_{2(x)} \mu_x + \beta_{1(x)}}{\beta_{2(x)} \bar{x} + \beta_{1(x)}} \right)$$

$$MSE(t_6) = \theta \mu_y^2 (C_y^2 + \theta_6^2 C_x^2 - 2\theta_6 \rho_{xy} C_x C_y) \quad (6)$$

where $\theta_5 = \frac{\beta_{2(x)} \mu_x}{\beta_{2(x)} \mu_x + \beta_{1(x)}}$

$$t_7 = \bar{y} \left(\frac{\beta_{1(x)} \mu_x + \beta_{2(x)}}{\beta_{1(x)} \bar{x} + \beta_{2(x)}} \right)$$

$$MSE(t_7) = \theta \mu_y^2 (C_y^2 + \theta_7^2 C_x^2 - 2\theta_7 \rho_{xy} C_x C_y) \quad (7)$$

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:

$$\theta = \frac{N-n}{Nn} = \frac{1-f}{n}, f = \frac{n}{N}, C_y = \frac{S_y}{\mu_y}, C_x = \frac{S_x}{\mu_x}$$

$$s_y^2 = (N-1)^{-1} \sum_{i=1}^N (y_i - \mu_y)^2$$

$$s_{xy}^2 = (N-1)^{-1} \sum_{i=1}^N (y_i - \mu_y)^2, \rho_{xy} = \frac{S_{xy}}{S_x S_y}$$

$$S_{xy} = (N-1)^{-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y).$$

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:

$$\hat{\mu}_y^R = t_8 = \frac{1}{n} \sum_{i=1}^n w_{yi}$$

The variance of $\hat{\mu}_y^R = \theta \mu_{wy}^2 C_{wy}^2$ (8)

where $\theta = \frac{N-n}{Nn} = \frac{1-f}{n}, f = \frac{n}{N}$ and $C_{wy}^2 = \frac{s_{wy}^2}{\mu_{wy}^2}$

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

$$t_9 = \bar{w}_y \left(\frac{\mu_x + C_{wx}}{\bar{w}_x + C_{wx}} \right)$$

$$t_{10} = \bar{w}_y \left(\frac{\mu_x \beta_{2(wx)} + C_{wx}}{\bar{w}_x \beta_{2(wx)} + C_{wx}} \right)$$

$$t_{11} = \bar{w}_y \left(\frac{\mu_x + \beta_{2(wx)}}{\bar{w}_x + \beta_{2(wx)}} \right)$$

The mean square error of $(t_9), (t_{10}), (t_{11})$, is given by:

$$MSE(t_9) = \theta \mu_{wy}^2 (C_{wy}^2 + \theta_{w7}^2 C_{wx}^2 - 2\theta_{w7} \rho_{wxwy} C_{wx} C_{wy}) \quad (9)$$

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

$$MSE(t_{10}) = \theta \mu_{wy}^2 (C_{wy}^2 + \theta_{w8}^2 C_{wx}^2 - 2\theta_{w8} \rho_{wxwy} C_{wx} C_{wy}) \quad (10)$$

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

$$MSE(t_{11}) = \theta \mu_{wy}^2 (C_{wy}^2 + \theta_{w9}^2 C_{wx}^2 - 2\theta_{w9} \rho_{wxwy} C_{wx} C_{wy}) \quad (11)$$

Where $\theta_{w9} = \frac{\mu_x}{\mu_x + \beta_{2wx}}$

Let the usual finite population consists N distinct units labelled from 1,2,..., N. the variables y_i and x_i ($i=1,2,3,\dots,N$) 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 ϕ_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 $\mu_{wy}, \mu_{wx}, \sigma_{wy}, \sigma_{wx}, C_{wy}, C_{wx}, \rho_{wxwy}, \sigma_{wywx}$ respectively.

The parameters of study and auxiliary variables are as:

Study variable

$$\mu_y = N^{-1} \sum_{i=1}^N y_i$$

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

Auxiliary variable

$$\mu_x = N^{-1} \sum_{i=1}^N x_i$$

$$\sigma_x^2 = (N - 1)^{-1} \sum_{i=1}^N (x_i - \mu_x)^2$$

Let $w_{yi} = \frac{1}{m_i} \sum_{j \in \phi_i} y_j$ and $w_{xi} = \frac{1}{m_i} \sum_{j \in \phi_i} x_j$ be the transformed survey and auxiliary variables in the i th network respectively.

The transformed population parameters of study and auxiliary variables are:

Study variable

$$\mu_{wy} = N^{-1} \sum_{i=1}^N w_{yi}$$

$$\sigma_{wy}^2 = (N - 1)^{-1} \sum_{i=1}^N (w_{yi} - \mu_{wy})^2$$

Auxiliary variable

$$\mu_{wx} = N^{-1} \sum_{i=1}^N w_{xi}$$

$$\sigma_{wx}^2 = (N - 1)^{-1} \sum_{i=1}^N (w_{xi} - \mu_{wx})^2$$

The transformed population statistics of study and auxiliary are:

Study variable

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

$$s_{wy}^2 = (n - 1)^{-1} \sum_{i=1}^n (w_{yi} - \bar{w}_y)^2$$

Auxiliary variable

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

$$s_{wx}^2 = (n - 1)^{-1} \sum_{i=1}^n (w_{xi} - \bar{w}_x)^2$$

III Results and Discussion

The proposed the generalized class of ratio type estimator under ACS as:

$$\hat{\mu}_y^r = t'_{p,q} = \bar{w}_y \left(\frac{p\beta_1(wx)\mu_x + \beta_2(wx)}{q\beta_1(wx)\bar{w}_x + \beta_2(wx)} \right),$$

where p and q are the constants to be determined in order to improve the efficiency of the proposed estimator.

The Mean square error of $\hat{\mu}_y^r = t'_{p,q}$ is given as:

$$\begin{aligned} \text{MSE}(\hat{\mu}_y^r) = t'_{p,q} = \mu_y^2 & \left[(\theta_{w10} - 1)^2 + \theta_{w10}^2 \theta C_{wy}^2 + \right. \\ & \theta_{w10} \theta_{w11}^2 \theta C_{wx}^2 (3\theta_{w13} - 2) + \\ & \left. \theta_{w10} \theta_{w11} \theta \rho_{wxwy} C_{wx} C_{wy} (2 - 4\theta_{w10}) \right] \end{aligned} \quad (12)$$

$$\text{Where } \theta_{w10} = \frac{p\beta_1(wx)\mu_x + \beta_2(wx)}{q\beta_1(wx)\mu_x + \beta_2(wx)}$$

$$\theta_{w11} = \frac{q\beta_1(wx)\mu_x}{q\beta_1(wx)\mu_x + \beta_2(wx)}$$

Special cases:

The proposed generalized estimator is an alternative to different types of estimators as already existing in literature, shown through different cases as follows;

- (i) For $p = q = 1$ the proposed generalized estimator becomes equal to $\bar{w}_y \left(\frac{\beta_1(wx)\mu_x + \beta_2(wx)}{\beta_1(wx)\bar{w}_x + \beta_2(wx)} \right)$ which is same as the estimator developed by S. K. Yadav et al. (2016). So it is found that the estimator proposed by S. K. Yadav et al. (2016) is a special case of the proposed generalized estimator.
- (ii) For $p = q = \beta_1(wx) = 1$, and $\beta_2(wx) = 0$ the proposed generalized estimator reduces to $\bar{w}_y \left(\frac{\mu_x}{\bar{w}_x} \right)$, which is same as the estimator developed by Dryver and Chao, (2007). So it is found that the estimator proposed by Dryver and Chao, (2007) is a special case of the proposed generalized estimator.

IV Numerical Study

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 I, the MSE values of proposed estimators for different values of p and q are given in table II, and percentage relative efficiency is given in table III.

Table I. Summary Statistics

$\mu_{wy}=1.2$ 225	$\mu_{wx}=0.5$ 50	$\bar{w}_y=1.22$ 25	$\bar{w}_x=0.55$ 0
$\beta_{1(wx)}=7$.953	$\beta_{2(wx)}=9$ 1.369	$C_{wx}=3.5$ 10	$C_{wy}=2.9$ 14
$S_{wx}=1.9$ 48	$S_{wy}=3.5$ 62	$S_{wxwy}=6$.428	$\rho_{wxwy}=0$.926
$\rho_{xy}=0.91$ 0	$C_x=4.32$ 5	$C_y=4.13$ 1	$\theta_1=0.11$ 4
$\theta_2=0.04$ 2	$\theta_3=0.06$ 4	$\theta_{w4}=0.0$ 06	$\theta_{w6}=0.0$ 46
$\beta_{1x}=608$ 32	$\beta_{2x}=55.0$ 90	N=400	n=20

$t'_{3,4}$	0.9222	0.1946	0.341
$t'_{3,8}$	0.8258	0.2787	0.282
$t'_{4,3}$	1.0422	0.1266	0.481
$t'_{4,4}$	1.0000	0.1619	0.408
$t'_{4,6}$	0.9251	0.2247	0.318
$t'_{5,4}$	1.0405	0.1619	0.437
$t'_{5,10}$	0.8371	0.3257	0.255

Table II.MSE of proposed estimator $t'_{p,q}$ at different values of p and q

$t'_{p,q}$	θ_{w10}	θ_{w11}	MSE
$t'_{1,1}$	1.0000	0.0461	0.543
$t'_{1,2}$	0.9559	0.0881	0.456
$t'_{1,3}$	0.9156	0.1266	0.396
$t'_{1,4}$	0.8785	0.1619	0.355
$t'_{1,5}$	0.8444	0.1946	0.329
$t'_{1,6}$	0.8127	0.2247	0.313
$t'_{1,7}$	0.7834	0.2527	0.304
$t'_{2,6}$	1.0000	0.0881	0.491
$t'_{2,8}$	0.9578	0.1266	0.418
$t'_{2,7}$	0.9190	0.1619	0.367
$t'_{2,1}$	0.7909	0.2787	0.290
$t'_{3,2}$	1.0000	0.1266	0.447
$t'_{3,3}$	0.9595	0.1619	0.385

From table II it has been found that the MSE of the proposed estimator shows irregular trend on increasing the values of p and q. Among these the MSE is least at p = 5 and q = 10. Therefore the proposed estimator at p = 5, q = 10 gives the optimum value of MSE and is therefore considered most efficient than Sisodia and Dwivedi (1981) Upadhyay and Singh (1999), Singh and Tailor (2003), Kadilar and Cingi (2003), Yan and Tian (2010), Thompson (1990), Chutiman (2013).

Table III. Percentage relative efficiency (PRE) of the proposed estimator with respect to existing estimators (EE), where the MSE of proposed estimator (PE) = 0.255.

Existing Estimator	MSE	PRE
$t_1 = \bar{y} \left(\frac{\mu_x + C_x}{\bar{x} + C_x} \right)$	0.966	378.823
$t_2 = \bar{y} \left(\frac{\beta_{2(x)}\mu_x + C_x}{\beta_{2(x)}\bar{x} + C_x} \right)$	0.407	159.607
$t_3 = \bar{y} \left(\frac{C_x\mu_x + \beta_{2(x)}}{C_x\bar{x} + \beta_{2(x)}} \right)$	1.117	438.039
$t_4 = \bar{y} \left(\frac{\mu_x + \rho_{xy}}{\bar{x} + \rho_{xy}} \right)$	0.527	206.666
$t_5 = \bar{y} \frac{\mu_x^2}{\bar{x}^2}$	1.904	746.666
$t_6 = \bar{y} \left(\frac{\beta_{2(x)}\mu_x + \beta_{1(x)}}{\beta_{2(x)}\bar{x} + \beta_{1(x)}} \right)$	0.311	121.960

V 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 MSE. Among the different cases developed on the base of proposed estimator the estimator $t'_{5,10}$ is best and should be preferred for the estimation of population mean when population is rare and clustered.

REFERENCES

Chutiman, N. (2013). Adaptive cluster sampling using auxiliary variable. *Journal of Mathematics and Statistics*, 9(3), 249-255.

Cochran, W. G. (1940). The estimation of yield of cereal experiment by sampling for the ratio of grain to total produce. *Journal of Agricultural Science*, 30, 262-1940.

Dryver, A. L., & Chao, C. T. (2007). Ratio estimators in adaptive cluster sampling. *Environmetrics*, 18, 607-620.

Hansen, M. H., & Hurwitz, W. N. (1943). On the theory of sampling from the finite population. *The Annals of Math Stat*, 14(4), 333-362.

Horvitz, D. G., & Thompson, D. J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, 47(260), 663-685.

Hussain, S., Sharma, M., & Jeelani, M. I. (2021). Optimum exponential ratio type estimator for the estimation of population mean. *Journal of Statistics Applications and Probability Letters*, 8(2), 73-82.

Kadilar, C., & Cingi, H. (2003a). A study on the chain ratio-type estimator. *Hacettepe Journal of Mathematics and Statistics*, 32, 105-108.

Kadilar, C., & Cingi, H. (2003b). Ratio estimators in stratified random sampling. *Biometrical Journal*, 45, 218-225.

Kumar, B., Sharma, M., Rizvi, S. E. H., & Jeelani, M. I. (2018). Estimation of population mean through improved class of ratio type estimator using auxiliary information. *Journal of Reliability and Statistical Studies*, 11(01), 01-07.

Robson, D. S. (1957). Applications of multivariate polykeys of the theory of unbiased ratio type estimation. *Journal American Statistical Association*, 52, 511-522.

Sharma, M., & Bhatnagar, S. (2008). A general class of estimator for the estimation of population mean using auxiliary information. *Journal of the Indian Statistical Association*, 46(2), 155-161.

Singh, R., & Mishra, R. (2022). Improved exponential ratio estimators in adaptive cluster sampling. *J. Stat. Appl. Pro. Lett.*, 11(1), 19-29.

$t_7 = \bar{y} \left(\frac{\beta_{1(x)}\mu_x + \beta_{2(x)}}{\beta_{1(x)}\bar{x} + \beta_{2(x)}} \right)$	1.068	418.823
$t_8 = \frac{1}{n} \sum_{i=1}^n w_{yi}$	0.603	236.470
$t_9 = \bar{w}_y \left(\frac{\mu_x + C_{wx}}{\bar{w}_x + C_{wx}} \right)$	0.432	169.411
$t_{10} = \bar{w}_y \left(\frac{\mu_x \beta_{2(wx)} + C_{wx}}{\bar{w}_x \beta_{2(wx)} + C_{wx}} \right)$	0.309	121.176
$t_{11} = \bar{w}_y \left(\frac{\mu_x + \beta_{2(wx)}}{\bar{w}_x + \beta_{2(wx)}} \right)$	0.595	233.333

From the above table it is found that at the optimum value (0.225) reveals that the proposed estimator is most efficient than existing estimators proposed by Sisodia and Dwivedi (1981) Upadhyay and Singh (1999), Singh and Tailor (2003), Kadilar and Cingi (2003), Yan and Tian (2010), Thompson (1990), Chutiman (2013). The percentage relative efficiency is shown with the help of bar diagram shown below:

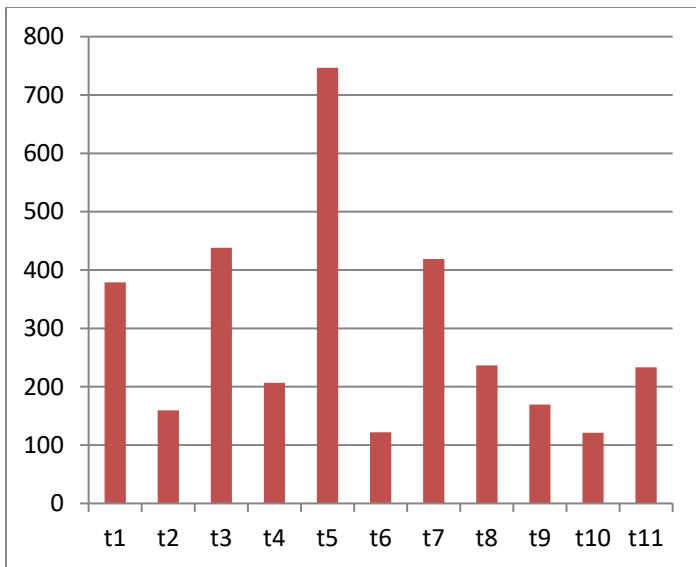


Fig. 2. PRE of various ratio type estimators w.r.t $t_{p,q}^{r2}$

The figure given above gives us the bird's eye view about the efficiency comparison of the proposed estimator while considering the optimum value of the proposed estimator. It reveals that the proposed estimator is most efficient than t_5 and less efficient than t_6, t_{10} in comparison to other existing estimators taken into literature. Thus the proposed estimator $t'_{5,10}$ is best to obtain the precise estimate of the population mean in rare and clustered population when there is positive correlation between the variables at the unit level.

- Singh, R., & Mishra, R. (2023). Wider class of estimators in ACS. *AJPS*, 24(2), 52-66.
- Sisodia, B. V. S., & Dwivedi, V. K. A. (1981). Modified ratio estimator using coefficient of variation of auxiliary variable. *Journal-Indian Society of Agricultural Statistics*, 33, 13-18.
- Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of American Statistical Association*, 85, 1050–1059.
- Upadhyay, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical journal*, 41(5), 627-636.
- Yan, Z., & Tian, B. (2010). Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable. In *International Conference on Information Computing and Applications*, 106, 103-110