AN EFFICIENT CLASS OF RATIO ESTIMATORS OF FINITE POPULATION MEAN UNDER SIMPLE RANDOM SAMPLING

[1]Mursala Khan, [2*]Sultan Hussain
[1,2] Department of Mathematics, COMSATS Institute of Information Technology, Abbottabad, 22060, Pakistan
mursala.khan@yahoo.com, tausef775650@yahoo.co.in

ABSTRACT*. In the present article, we have proposed a class of ratio estimators for the estimation of finite population mean using simple random sampling scheme when there is maximum and minimum values on both the study and the auxiliary variables and their properties are considered up to the first order of approximation. Also, we have found some theoretical conditions which make the proposed class of ratio estimators more efficient than the Kadilar and Cingi [1] estimators and other relevant existing estimators. In addition, these theoretical results are supported by a real data set.*

Key Words: Population mean, median, study variate, ratio estimators, maximum and minimum values, bias, mean square error, simple random sampling.

1. INTRODUCTION

In the field of survey sampling, the use of auxiliary information plays an important role in the estimation of population mean of the variable under study both the design as well as the estimation stage and have greatly increase the efficiency of the estimators. Various statisticians have worked on the estimation of the unknown population mean of the study variable using auxiliary information. Using auxiliary information Sisodia and Dwivedi [2] proposed ratio estimator using the known knowledge coefficient of variation of an auxiliary variable, Kadilar and Cingi [1] suggested a class of ratio estimators for the finite population mean, including Upadhyaya and Singh [3] , Singh [4], Singh and Tailor [5], Singh et al. [6], Kadilar and Cingi [7, 8], khan [9], Khan and Hussain [10], Khan et al. [11] and Yan and Tian [12] and Yadav and Kadilar [13] have used auxiliary information for improved estimation of population mean of variable under study *y*.

Consider a finite population of size *N* units such that $U=\{U_1, U_2, U_3...., U_N\}$. Let *y* and *x* be the study and the auxiliary variable with corresponding values y_i and x_i respectively for the *i*-th unit $i=\{1, 2, 3...., N\}$ defined on a finite population *U*.

Let $\bar{Y}=(1/N)\sum_{i=1}^{N} y_i$ and $\bar{X}=(1/N)\sum_{i=1}^{N} x_i$ be the population means of the study and the auxiliary variable, respectively.

And $S_y^2=(1/N-1)\sum_{i=1}^{N}(y_i-\bar{Y})^2$ and $S_x^2=(1/N-1)\sum_{i=1}^{N}(x_i-\bar{X})^2$ be the corresponding population mean square error of the study and the auxiliary variable, respectively and let $C_y=\dfrac{S_y}{\bar{Y}}$ and $C_x=\dfrac{S_x}{\bar{X}}$

be the coefficient of variation of the study as well as auxiliary variable respectively, and $\rho_{yx}=\dfrac{S_{yx}}{S_y S_x}$ be the population correlation coefficient between *x* and *y*.

In order to estimate the unknown population parameters we take a random sample of size *n* units from the

finite population *U* by using simple random sample without replacement.

Let $\bar{y}=(1/n)\sum_{i=1}^{n} y_i$ and $\bar{x}=(1/n)\sum_{i=1}^{n} x_i$ be the corresponding

sample means of the study and the auxiliary variable, respectively, and their corresponding sample variances are

$\hat{S}_y^2=(1/n-1)\sum_{i=1}^{n}(y_i-\bar{y})^2$ and $\hat{S}_x^2=(1/n-1)\sum_{i=1}^{n}(x_i-\bar{x})^2$

respectively.

The usual unbiased estimator for population mean $\bar{Y}$ of the study variable is, given by

$$\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n} \tag{1.1}$$

The variance of the estimator $\bar{y}$ up to first order of approximation is, given by

$$Var(\bar{y})=\lambda S_y^2 \tag{1.2}$$

Where $\lambda=\dfrac{1}{n}-\dfrac{1}{N}$,

In many practical situations, when we want to estimate the unknown characteristic of the population there exists some large ($y_{\max}$) or small values ($y_{\min}$) and to estimate the population quantities without considering these information is very sensitive in either the case the result will be either over estimated or under estimated. In order to tackle this situation Sarndal [14], proposed the following unbiased estimator for finite population mean.

$$\bar{y}_S = \begin{cases} \bar{y}+c \text{ if sample contains } y_{\min} \text{ but not } y_{\max} \\ \bar{y}-c \text{ if sample contains } y_{\max} \text{ but not } y_{\min} \\ \bar{y} \quad \text{for all other samples,} \end{cases} \quad (1.3)$$

where c is a constant, whose values is to be find for minimum variance

The minimum variance of the estimator $\bar{y}_S$ up to first order of approximation is, given as

$$\operatorname{var}\left(\bar{y}_S\right)_{\min} = \operatorname{var}\left(\bar{y}\right) - \frac{\lambda\left(y_{\max}-y_{\min}\right)^2}{2\left(N-1\right)} \quad (1.4)$$

where the optimum value of c_{opt} is $c_{opt} = \dfrac{\left(y_{\max}-y_{\min}\right)}{2n}$.

Using the known knowledge of the auxiliary variables Kadilar and Cingi [1], suggested the following class of ratio estimators for population mean of the variable under study under simple random sampling $\bar{y}_{kci}$ and is given by

$$\hat{\bar{Y}}_{KC1} = \frac{\bar{y}+b_{yx}\left(\bar{X}-\bar{x}\right)}{\bar{x}}\bar{X} \quad (1.5)$$

$$\hat{\bar{Y}}_{KC2} = \frac{\bar{y}+b_{yx}\left(\bar{X}-\bar{x}\right)}{\left(\bar{x}+C_x\right)}\left(\bar{X}+C_x\right) \quad (1.6)$$

$$\hat{\bar{Y}}_{KC3} = \frac{\bar{y}+b_{yx}\left(\bar{X}-\bar{x}\right)}{\left(\bar{x}+M_x\right)}\left(\bar{X}+\beta_{2x}\right) \quad (1.7)$$

$$\hat{\bar{Y}}_{KC4} = \frac{\bar{y}+b_{yx}\left(\bar{X}-\bar{x}\right)}{\left(\bar{x}\beta_{2x}+C_x\right)}\left(\bar{X}\beta_{2x}+C_x\right) \quad (1.8)$$

$$\hat{\bar{Y}}_{KC5} = \frac{\bar{y}+b_{yx}\left(\bar{X}-\bar{x}\right)}{\left(\bar{x}C_x+\beta_{2x}\right)}\left(\bar{X}C_x+\beta_{2x}\right) \quad (1.9)$$

where β_{2x} is the known value of the population coefficient of kurtosis of the auxiliary variable,

The bias and mean square error of the estimators $\bar{y}_{kci}$ up to first order of approximation are, given by

$$Bias\left(\hat{\bar{Y}}_{KCi}\right) = \frac{\alpha_{KCi}\lambda}{\bar{X}}\left[\left(\frac{\alpha_{KCi}}{2}+\beta_{yx}\right)S_x^2 - S_{yx}\right], \quad (1.10)$$

$$MSE\left(\hat{\bar{Y}}_{KCi}\right) = \lambda\left[\alpha_{KCi}^2 S_x^2 + S_y^2\left(1-\rho_{yx}^2\right)\right], \quad (1.11)$$

for $i = 1, 2, 3, 4, 5.$

where $\alpha_{KC1} = \dfrac{\bar{Y}}{\bar{X}}$, $\alpha_{KC2} = \dfrac{\bar{Y}}{\bar{X}+C_x}$, $\alpha_{KC3} = \dfrac{\bar{Y}}{\bar{X}+\beta_{2x}}$,

$\alpha_{KC4} = \dfrac{\bar{Y}\beta_{2x}}{\bar{X}\beta_{2x}+C_x}$ and $\alpha_{KC5} = \dfrac{\bar{Y}C_x}{\bar{X}C_x+\beta_{2x}}$.

2. The Proposed Class of Ratio Estimators

On the lines of Sarndal [14], we proposed a class of ratio estimators for the estimation of finite population mean of the variable under study *y*, using the known knowledge of an auxiliary variable say *x*. Usually when the relationship between the study and the auxiliary variable is positive then the selection of the larger value of the auxiliary variable the larger the value of the study variable is to be expected, and the selection of the smaller value of the auxiliary variable the smaller the value of the variable under study is to be expected. Using the above information we proposed the following class of ratio-type estimators.

$$\hat{\bar{Y}}_{M1} = \frac{\bar{y}_{C_{11}}+b_{yx}\left(\bar{X}-\bar{x}_{C_{21}}\right)}{\bar{x}_{C_{21}}}\bar{X} \quad (2.1)$$

$$\hat{\bar{Y}}_{M2} = \frac{\bar{y}_{C_{11}}+b_{yx}\left(\bar{X}-\bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}}+C_x\right)}\left(\bar{X}+C_x\right) \quad (2.2)$$

$$\hat{\bar{Y}}_{M3} = \frac{\bar{y}_{C_{11}}+b_{yx}\left(\bar{X}-\bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}}+M_x\right)}\left(\bar{X}+\beta_{2x}\right) \quad (2.3)$$

$$\hat{\bar{Y}}_{M4} = \frac{\bar{y}_{C_{11}}+b_{yx}\left(\bar{X}-\bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}}\beta_{2x}+C_x\right)}\left(\bar{X}\beta_{2x}+C_x\right) \quad (2.4)$$

$$\hat{\bar{Y}}_{M5} = \frac{\bar{y}_{C_{11}}+b_{yx}\left(\bar{X}-\bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}}C_x+\beta_{2x}\right)}\left(\bar{X}C_x+\beta_{2x}\right) \quad (2.5)$$

$$\hat{\bar{Y}}_{M6} = \frac{\bar{y}_{C_{11}}+b_{yx}\left(\bar{X}-\bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}}+M_x\right)}\left(\bar{X}+M_x\right) \quad (2.6)$$

$$\hat{\bar{Y}}_{M7} = \frac{\bar{y}_{C_{11}} + b_{yx}\left(\bar{X} - \bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}} M_x + \rho_{yx}\right)}\left(\bar{X} M_x + \rho_{yx}\right) \qquad (2.7)$$

$$\hat{\bar{Y}}_{M8} = \frac{\bar{y}_{C_{11}} + b_{yx}\left(\bar{X} - \bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}} \rho_{yx} + M_x\right)}\left(\bar{X} \rho_{yx} + M_x\right) \qquad (2.8)$$

$$\hat{\bar{Y}}_{M9} = \frac{\bar{y}_{C_{11}} + b_{yx}\left(\bar{X} - \bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}} C_x + M_x\right)}\left(\bar{X} C_x + M_x\right) \qquad (2.9)$$

$$\hat{\bar{Y}}_{M10} = \frac{\bar{y}_{C_{11}} + b_{yx}\left(\bar{X} - \bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}} \beta_{2x} + M_x\right)}\left(\bar{X} \beta_{2x} + M_x\right) \qquad (2.10)$$

$$\hat{\bar{Y}}_{M11} = \frac{\bar{y}_{C_{11}} + b_{yx}\left(\bar{X} - \bar{x}_{C_{21}}\right)}{\left(\bar{x}_{C_{21}} M_x + \beta_{2x}\right)}\left(\bar{X} M_x + \beta_{2x}\right) \qquad (2.11)$$

where M_x and β_{2x} are the known values of the population median and population coefficient of kurtosis of the auxiliary variable respectively and $\left(\bar{y}_{C_{11}} = \bar{y} + c_1, \quad \bar{x}_{C_{21}} = \bar{x} + c_2\right)$, where c_1 and c_2 are unknown constants whose values are to be find for optimality conditions.

To obtain the properties of $\hat{\bar{Y}}_{Mi}$ in terms of bias and Mean square error, we define the following relative error terms and their expectations.

$\zeta_0 = \frac{\bar{y}_{C_1} - \bar{Y}}{\bar{Y}}, \quad \zeta_1 = \frac{\bar{x}_{C_2} - \bar{X}}{\bar{X}}$, such that

$E(\zeta_0) = E(\zeta_1) = 0$, also

$$E\left(\zeta_0^2\right) = \frac{\theta}{\bar{Y}^2}\left(S_y^2 - \frac{2nc_1}{N-1}\left(y_{\max} - y_{\min} - nc_1\right)\right),$$

$$E\left(\zeta_1^2\right) = \frac{\theta}{\bar{X}^2}\left(S_x^2 - \frac{2nc_2}{N-1}\left(x_{\max} - x_{\min} - nc_2\right)\right) \text{ and}$$

$$E\left(\zeta_0\zeta_1\right) = \frac{\theta}{\overline{YX}}\left(S_{yx} - \frac{n}{N-1}\begin{pmatrix} c_2\left(y_{\max} - y_{\min}\right) + \\ c_1\left(x_{\max} - x_{\min}\right) - 2nc_1c_2 \end{pmatrix}\right).$$

The biases and mean square errors of the estimators $\hat{\bar{Y}}_{Mi}$ up to first order of approximation are, given by

$$Bias\left(\hat{\bar{Y}}_{Mj}\right) =$$

$$\frac{\alpha_{Mj}\lambda}{\bar{X}}\left[\begin{array}{l} \left(\frac{\alpha_{Mj}}{2} + \beta_{yx}\right)\left(S_x^2 - \frac{2nc_2}{N-1}\begin{pmatrix} x_{\max} - x_{\min} \\ -nc_2 \end{pmatrix}\right) - S_{yx} \\ + \frac{n}{N-1}\begin{pmatrix} c_2\left(y_{\max} - y_{\min}\right) + c_1\left(x_{\max} - x_{\min}\right) \\ -2nc_1c_2 \end{pmatrix} \end{array}\right] \qquad (2.12)$$

and

$$MSE\left(\hat{\bar{Y}}_{Mj}\right)_{\min} =$$

$$\lambda\left[\begin{array}{l} \alpha_{Mj}^2 S_x^2 + S_y^2\left(1 - \rho_{yx}^2\right) \\ - \frac{1}{2(N-1)}\begin{pmatrix} \left(y_{\max} - y_{\min}\right) - \\ \left(\alpha_{Mj} + \beta_{yx}\right)\left(x_{\max} - x_{\min}\right) \end{pmatrix}^2 \end{array}\right] \qquad (2.13)$$

for $j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11.$

where the optimum values of c_1 and c_2 are given by

$$\left.\begin{array}{l} c_1 = \frac{\left(y_{\max} - y_{\min}\right)}{2n} \\ c_2 = \frac{\left(x_{\max} - x_{\min}\right)}{2n} \end{array}\right\} \qquad (2.14)$$

where $\alpha_{M1} = \frac{\bar{Y}}{\bar{X}}, \alpha_{M2} = \frac{\bar{Y}}{\bar{X} + C_x}, \alpha_{M3} = \frac{\bar{Y}}{\bar{X} + \beta_{2x}},$

$\alpha_{M4} = \frac{\bar{Y}\beta_{2x}}{\bar{X}\beta_{2x} + C_x}, \alpha_{M5} = \frac{\bar{Y}C_x}{\bar{X}C_x + \beta_{2x}}$

$\alpha_{M6} = \frac{\bar{Y}}{\bar{X} + M_x}, \alpha_{M7} = \frac{\bar{Y}M_x}{\bar{X}M_x + \rho_{yx}},$

$\alpha_{M8} = \frac{\bar{Y}\rho_{yx}}{\bar{X}\rho_{yx} + M_x}, \alpha_{M9} = \frac{\bar{Y}C_x}{\bar{X}C_x + M_x},$

$\alpha_{M10} = \frac{\bar{Y}\beta_{2x}}{\bar{X}\beta_{2x} + M_x}$ and $\alpha_{M11} = \frac{\bar{Y}M_x}{\bar{X}M_x + \beta_{2x}}.$

3. Comparison of Estimators

In this section, we have made some theoretical conditions under which our proposed class of ratio estimators performs better than the other relevant existing estimators discussed in the literature of survey sampling.

(i) By (1.11) and (2.13),

$$\left[MSE\left(\hat{\bar{Y}}_{KCi}\right)-MSE\left(\hat{\bar{Y}}_{Mi}\right)_{\min}\right]\geq 0 \text{ if,}$$

$$\left[\left(\alpha_{KCi}^2-\alpha_{Mj}^2\right)S_x^2+\frac{1}{2(N-1)}\left\{\begin{matrix}\left(y_{\max}-y_{\min}\right)- \\ \left(\alpha_{Mj}+\beta_{yx}\right)\left(x_{\max}-x_{\min}\right)\end{matrix}\right\}^2\right]\geq 0.$$

for $i=1,2,3,4,5.$ and $j=1,2,3,4,5,6,7,8,9,10,11.$

4 Empirical Study

To examine the performance of the proposed class of ratio estimators with some of the existing estimators discussed in the literature, we have considered a natural population from the literature. The description and the necessary data statistics of the population are given below.

Population. (Source: Agricultural Statistics (1999) [15], Washington, US.).

Y : Amount (in \$000) of real estate farms loans in different estates during 1997.

X : Amount (in \$000) of non-real estate farms loans in different estates during 1997.

$$N=50,\ n=15,\ \bar{Y}=555.43,\ \bar{X}=878.16,\ S_y^2=342021.5,$$
$$S_x^2=1176526,\ \rho_{yx}=0.804,\ S_{yx}=509910.41,\ C_y^2=1.109,$$
$$C_x^2=1.526,\ Y_{\max}=2327.03,\ Y_{\min}=1.61,\ X_{\max}=3928.732,$$
$$X_{\min}=0.233,\ M_x=452.517,\ \beta_{2x}=4.5247,\ \beta_{yx}=0.4334.$$

The mean squared errors of the existing and the proposed class estimators are shown in **Table-1.**

Table 1. *MSE's* of the competing and the proposed class of ratio estimators

Population		
Estimators		*MSE*'s
Existing	$\hat{\bar{Y}}_{KC1}$ Kadilar and Cingi [1]	27632.8282
	$\hat{\bar{Y}}_{KC2}$ Kadilar and Cingi [1]	27571.1249
	$\hat{\bar{Y}}_{KC3}$ Kadilar and Cingi [1]	27408.0117
	$\hat{\bar{Y}}_{KC4}$ Kadilar and Cingi [1]	27619.4203
	$\hat{\bar{Y}}_{KC5}$ Kadilar and Cingi [1]	27450.5951
Proposed	$\hat{\bar{Y}}_{M1}$ Proposed Estimator	25984.4793
	$\hat{\bar{Y}}_{M2}$ Proposed Estimator	25928.2290
	$\hat{\bar{Y}}_{M3}$ Proposed Estimator	25781.7167
	$\hat{\bar{Y}}_{M4}$ Proposed Estimator	27192.5264
	$\hat{\bar{Y}}_{M5}$ Proposed Estimator	25820.3199
	$\hat{\bar{Y}}_{M6}$ Proposed Estimator	14732.4339
	$\hat{\bar{Y}}_{M7}$ Proposed Estimator	25984.2055
	$\hat{\bar{Y}}_{M8}$ Proposed Estimator	13436.6316
	$\hat{\bar{Y}}_{M9}$ Proposed Estimator	15989.51339
	$\hat{\bar{Y}}_{M10}$ Proposed Estimator	22138.8749
	$\hat{\bar{Y}}_{M11}$ Proposed Estimator	25984.4793

5. CONCLUSION

In this research article, we have proposed a class of ratio estimators using the known knowledge of the auxiliary variable such as population mean, population median, population coefficient of variation, population coefficient of kurtosis for finite population mean under maximum and minimum using simple random sampling. We have also find some theoretical conditions under which the suggested class of ratio-type estimators have always efficient than the Kadilar and Cingi [1] estimators. Theoretical results are also verified with the help of real data set which clearly indicates that the proposed class of ratio estimators have smaller mean square error than the other estimators discussed in the literature. Thus the suggested approach under maximum and minimum values may be preferred over the existing estimators for the use of applied applications.

REFERENCES

[1] C. Kadilar, H. Cingi, Improvement in estimating the population mean in simple random sampling, Applied Mathematics Letters, 19, (2006), 75-79.

[2] B. V. S. Sisodia, V. K. A. Dwivedi, Modified ratio estimator using coefficient of variation of auxiliary variable, Journal of the Indian Society of Agricultural Statistics, 33, (1981), 13-18.

[3] L. N. Upadhyaya, H. P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, Biometrical Journal, 41(5), (1999), 627-636.

[4] G. N. Singh, On the improvement of product method of estimation in sample surveys, Journal of the Indian Society of Agricultural Statistics, 56(3), (2003), 267–265.

[5] H. P. Singh, R. Tailor, Use of known correlation coefficient in estimating the finite population means, Statistics in Transition, 6(4), (2003), 555-560.

[6] H. P. Singh, R. Tailor, M. S. Kakran, An Improved Estimator of population mean using Power transformation, Journal of the Indian Society of Agricultural Statistics, 58(2), (2004), 223-230.

[7] C. Kadilar, H. Cingi, Ratio estimators in simple random sampling, Applied Mathematics and Computation, 151, (2004), 893-902.

[8] C. Kadilar, H. Cingi, An improvement in estimating the population mean by using the correlation coefficient, Hacettepe Journal of mathematics and Statistics, 35, (2006), 103-109.

[9] M. Khan, Improvement in estimating the finite population mean under maximum and minimum values in double sampling scheme, Journal of Statistics Applications & Probability Letters, 2(2), (2015), 1-7.

[10] M. Khan, S. Hussain, An improved class of ratio-type estimators for finite population mean under maximum and minimum values, Science International, 27(2), (2015), 1135-1138.

[11] M. Khan, S. Ullah, A. Hossain, N. Bashir, Improved ratio-type estimators using maximum and minimum values under simple random sampling scheme, Hacettepe Journal of Mathematics and Statistics, (2015), (accepted).

[12] Z, Yan, B, Tian, Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable, ICICA 2010, Part II, CCIS 106, 103–110.

[13] S. K. Yadav, C. Kadilar, Improved class of ratio and product estimators, Applied Mathematics and Computation, 219, (2013), 10726–10731.

[14] C. E. Sarndal, Sample survey theory vs general statistical theory: Estimation of the population mean, International Statistical Institute, 40, (1972), 1-12.

[15] http://www.nass.usda.gov/Publications/Ag_Statistics/199_9/index.asp.