Hacettepe Journal of Mathematics and Statistics
Volume 44 (4) (2015), 923 – 931

Improved ratio-type estimators using maximum and minimum values under simple random sampling scheme

Mursala Khan[*†], Saif Ullah[‡], Abdullah Y. Al-Hossain[§] and Neelam Bashir[¶]

Abstract

This paper presents a class of ratio-type estimators for the evaluation of finite population mean under maximum and minimum values by using knowledge of the auxiliary variable. The properties of the proposed estimators in terms of biases and mean square errors are derived up to first order of approximation. Also, the performance of the proposed class of estimators is shown theoretically and these theoretical conditions are, then, verified numerically by taking three natural populations under which the proposed class of estimators performed better than other competing estimators.

2000 AMS Classification: 62D05.

Keywords: Study variable, Auxiliary variable, Ratio estimators, Maximum and Minimum values, Simple random sampling, Mean squared error, Efficiency.

Received 18/04/2014 : Accepted 06/09/2014 Doi : 10.15672/HJMS.2014297480

---

[*] Department of Mathematics, COMSATS Institute of Information Technology Abbottabad, 22060, Pakistan. Email: `mursala.khan@yahoo.com`

[†] Corresponding Author.

[‡] Department of Mathematics, Government College University Lahore, 54000, Pakistan. Email: `dr.saifullah@gcu.edu.pk`

[§] Department of Mathematics, Faculty of Science, Jazan University, Jazan 2097, Saudi Arabia.

[¶] Department of Mathematics, Government College University Lahore, 54000, Pakistan.

1. Introduction

A lot of work has been done for the estimation of finite population mean using auxiliary information for improving the efficiency of the estimators. Das and Tripathi (1980, 1981), Uphadhyaya and Singh (1999), Singh (2004), Sisodia and Dwivedi (1981) proposed ratio estimator using coefficient of variation of an auxiliary variable. Kadilar and Cingi (2005) suggested ratio estimators in stratified random sampling. In the same way, Kadilar and Cingi (2006) proposed an improvement in estimating the population mean by using the correlation coefficient. Khan and Shabbir (2013) proposed a ratio-type estimator for the estimation of population variance using the knowledge of quartiles and their functions as auxiliary information. They proposed different modified estimators for the estimation of finite population mean using maximum and minimum values. Recently Hossain and Khan (2014) worked on the estimation of population mean using maximum and minimum values under simple random sampling by incorporating the knowledge of two auxiliary variables.

Let us consider a finite population of size N of different units $U = \{U_1, U_2, U_3, ..., U_N\}$. 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 $\overline{Y} = (1/N) \sum_{i=1}^{N} y_i$ and $\overline{X} = (1/N) \sum_{i=1}^{N} x_i$ be the population means of the study and the auxiliary variable, respectively. Also $S_y^2 = (1/N-1) \sum_{i=1}^{N} (y_i - \overline{Y})^2$ and $S_x^2 = (1/N-1) \sum_{i=1}^{N} (x_i - \overline{X})^2$ be the corresponding population mean square error of the study and the auxiliary variable respectively, and let $C_y = S_y/\overline{Y}$ and $C_x = S_x/\overline{X}$ be the coefficients of variation of the study and the auxiliary variable respectively, and $\rho_{yx} = S_{yx}/S_y S_x$ 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 $\overline{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\overline{x} = (1/n) \sum_{i=1}^{n} x_i$ be the corresponding sample means of the study and the auxiliary variable respectively, and their corresponding sample variances are $\hat{S}_y^2 = (1/n-1) \sum_{i=1}^{n} (y_i - \overline{y})^2$ and $\hat{S}_x^2 = (1/n-1) \sum_{i=1}^{n} (x_i - \overline{x})^2$ respectively.

When there is no auxiliary information the usual unbiased estimator for the population mean of the study variable is:

$$(1.1) \qquad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

The variance of the estimator $\overline{y}$ is given by:

$$(1.2) \qquad \operatorname{var}(\overline{y}) = \theta\, S_y^2 \,, \quad \text{where } \theta = \frac{1}{n} - \frac{1}{N}.$$

In many populations there exist some large (y_{max}) or small (y_{min}) values and to estimate the population parameters without considering this information is very sensitive. In either case the result will be overestimated or underestimated. In order to handle this situation Sarndal (1972) suggested the following unbiased estimator for the assessment of finite population mean:

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

were c is a constant whose value is to be found for minimum variance.

The minimum variance of the estimator $\overline{y}_s$ up to first order of approximation is given as under:

$$\text{var}\,(\overline{y}_s)_{min} = \text{var}\,(\overline{y}) - \frac{\theta(y_{max} - y_{min})^2}{2(N-1)}\,, \tag{1.4}$$

where the optimum value of c_{opt} is

$$c_{opt} = \frac{(y_{max} - y_{min})}{2n}.$$

The classical ratio estimator for finding the population mean of the study variable is given by:

$$\hat{\overline{Y}}_R = \overline{y}\,\frac{\overline{X}}{\overline{x}} \tag{1.5}$$

The bias and mean square errors of the estimator $\hat{\overline{Y}}_R$ up to first order of approximation are given by:

$$\text{Bias}\left(\hat{\overline{Y}}_R\right) = \frac{\theta}{\overline{X}}\,(RS_x^2 - S_{yx}) \tag{1.6}$$

$$\text{MSE}\left(\hat{\overline{Y}}_R\right) = \theta(S_y^2 + R^2S_x^2 - 2RS_{yx}) \tag{1.7}$$

Similarly, Sisodid and Dwivedi (1981) suggested the following ratio estimator using the knowledge of coefficient of variation of the auxiliary variable:

$$\hat{\overline{Y}}_{SD} = \overline{y}\left(\frac{\overline{X} + C_x}{\overline{x} + C_x}\right) \tag{1.8}$$

The bias and mean square errors of the estimator $\hat{\overline{Y}}_{SD}$ up to first order of approximation are as follows:

$$\text{Bias}\left(\hat{\overline{Y}}_{SD}\right) = \frac{\theta\,\alpha_1}{\overline{Y}}\,(R\alpha_1S_x^2 - S_{yx}), \tag{1.9}$$

$$\text{MSE}\left(\hat{\overline{Y}}_{SD}\right) = \theta(S_y^2 + \alpha_1^2S_x^2 - 2\alpha_1S_{yx}), \quad \text{where } \alpha_1 = \frac{\overline{Y}}{\overline{X} + C_x}\,. \tag{1.10}$$

2. The proposed class of estimators

On the lines of Sarndal (1972), we propose a class of ratio-type estimators for the estimation of finite population mean using knowledge of the coefficient of variation and coefficient of correlation of an auxiliary variable. Usually when the correlation between the study variable (y) and the auxiliary variable (x) is positive, then the selection of the larger value of the auxiliary variable (x), the larger value of study variable (y) is to be expected, and the smaller the value of auxiliary variable (x), the smaller the value of study variable (y). Using such type of information, we propose the following class of estimators given by:

$$\hat{\overline{Y}}_{P_1} = \overline{y}_{c_1}\left(\frac{\overline{X} + C_x}{\overline{x}_{c_2} + C_x}\right), \tag{2.1}$$

$$\hat{\overline{Y}}_{P_2} = \overline{y}_{c_1}\left(\frac{\overline{X} + \rho_{yx}}{\overline{x}_{c_2} + \rho_{yx}}\right), \tag{2.2}$$

$$\hat{\overline{Y}}_{P_3} = \overline{y}_{c_1}\left(\frac{\overline{X}C_x + \rho_{yx}}{\overline{x}_{c_2}C_x + \rho_{yx}}\right), \tag{2.3}$$

$$\hat{\overline{Y}}_{P_4} = \overline{y}_{c_1}\Big(\frac{\overline{X}\rho_{yx} + C_x}{\overline{x}_{c_2}\rho_{yx} + C_x}\Big), \tag{2.4}$$

where $(\overline{y}_{c_1} = \overline{y} + c_1,\ \overline{x}_{c_2} = \overline{x} + c_2)$, also c_1 and c_2 are unknown constants.

To obtain the properties of $\hat{\overline{Y}}_{P_i}$ in terms of Bias and Mean square error, we define the following relative error terms and their expectations:

$$\zeta_0 = \frac{\overline{y}_{c_1} - \overline{Y}}{\overline{Y}}\ , \quad \zeta_1 = \frac{\overline{x}_{c_2} - \overline{X}}{\overline{X}}\ , \text{ such that } \mathrm{E}(\zeta_0) = \mathrm{E}(\zeta_1) = 0,$$

Also,

$$\mathrm{E}\Big(\zeta_0^2\Big) = \frac{\theta}{\overline{Y}^2}\Big(S_y^2 - \frac{2nc_1}{N-1}(y_{max} - y_{min} - nc_1)\Big)\ ,\ \ \mathrm{E}\Big(\zeta_1^2\Big) = \frac{\theta}{\overline{X}^2}\Big(S_x^2 - \frac{2nc_2}{N-1}(x_{max} - x_{min} - nc_2)\Big)$$

$$\text{and} \quad \mathrm{E}(\zeta_0\zeta_1) = \frac{\theta}{\overline{Y}\,\overline{X}}\Big(S_{yx} - \frac{n}{N-1}(c_2(y_{max} - y_{min}) + c_1(x_{max} - x_{min}) - 2nc_1c_2)\Big)\ .$$

$$\begin{aligned} \text{Where } \theta \quad &= \quad \frac{1}{n} - \frac{1}{N}\ ,\ R = \frac{\overline{Y}}{\overline{X}}\ ,\ \alpha_{P_1} = \frac{\overline{X}}{\overline{X} + C_x}\ ,\ \alpha_{P_2} = \frac{\overline{X}}{\overline{X} + \rho_{yx}}\ ,\ \alpha_{P_3} = \frac{\overline{X}C_x}{\overline{X}C_x + \rho_{yx}}\ , \\ & \alpha_{P_4} = \frac{\overline{X}\rho_{yx}}{\overline{X}\rho_{yx} + C_x}\ ,\ k_{P1} = \frac{\overline{Y}}{\overline{X} + C_x}\ ,\ k_{P2} = \frac{\overline{Y}}{\overline{X} + \rho_{yx}}\ ,\ k_{P3} = \frac{\overline{Y}C_x}{\overline{X}C_x + \rho_{yx}} \\ & \text{and} \quad k_{P4} = \frac{\overline{Y}\rho_{yx}}{\overline{X}\rho_{yx} + C_x}. \end{aligned}$$

Rewriting $\hat{\overline{Y}}_{P_i}$ in terms of ζ_i's, we have

$$\hat{\overline{Y}}_{P_i} = \overline{Y}\ \Big(1 + \zeta_0\Big)\Big(1 + \alpha_{P_i}\zeta_1\Big)^{-1},$$

where $\hat{\overline{Y}}_{P_i}$ represent the proposed class of estimators for i=1, 2, 3, 4.

Expanding the right hand side of the equation given above and including terms up to second powers of ζ_i's i.e., up to first order of approximation, we have:

$$\hat{\overline{Y}}_{P_i} - \overline{Y} = \overline{Y}\Big(\zeta_0 - \alpha_{P_i}\zeta_1 + \alpha_{Pi}^2\zeta_1^2 - \alpha_{P_i}\zeta_0\zeta_1\Big) \tag{2.5}$$

Taking expectation on both sides of (2.5), we get bias up to first order of approximation which is given as:

$$\begin{aligned} \mathrm{Bias}(\hat{\overline{Y}}_{P_i}) \quad &= \quad \frac{\theta\, k_{P_i}}{\overline{Y}}\Big[k_{P_i}\Big(S_x^2 - \frac{2nc_2}{N-1}(x_{max} - x_{min} - nc_2)\Big) - S_{yx} \\ & \quad + \frac{n}{N-1}(c_2(y_{max} - y_{min}) + c_1(x_{max} - x_{min}) - 2nc_1c_2)\Big] \end{aligned} \tag{2.6}$$

On squaring both sides of (2.5), and keeping ζ_i's powers up to first order of approximation, we get:

$$\Big(\hat{\overline{Y}}_{P_i} - \overline{Y}\Big)^2 = \overline{Y}^2\ \Big(\zeta_0^2 + \alpha_{P_i}^2\zeta_1^2 - 2\alpha_{P_i}\zeta_0\zeta_1\Big) \tag{2.7}$$

Taking expectation on both sides of (2.7), we get mean square error up to first order of approximation, given as under:

$$\begin{aligned} \mathrm{MSE}\Big(\hat{\overline{Y}}_{P_i}\Big) \quad &= \quad \theta\Big[\Big(S_y^2 + k_{P_i}^2S_x^2 - 2k_{P_i}S_{yx}\Big) - \frac{2n}{N-1}\Big\{(c_1 - c_2k_{P_i})((y_{max} - y_{min}) \\ & \quad - n(c_1 - c_2k_{P_i}) - k_{P_i}(x_{max} - x_{min}))\Big\}\Big] \end{aligned} \tag{2.8}$$

The optimum values of c_1 and c_2 are given in the following lines:

$$\left\{\begin{array}{l} c_1 = \frac{(y_{max}-y_{min})}{2n} \\ \\ c_2 = \frac{(x_{max}-x_{min})}{2n} \end{array}\right. \tag{2.9}$$

On substituting the optimum value of c_1 and c_2 in (2.8), we get the minimum mean square error of the proposed estimators as follows:

$$\begin{aligned} \text{MSE}\left(\hat{\bar{Y}}_{P_i}\right)_{min} &= \theta\Big[\Big(S_y^2 + k_{P_i}^2 S_x^2 - 2k_{P_i} S_{yx}\Big) \\ &\quad -\frac{1}{2(N-1)}\Big((y_{max}-y_{min}) - k_{P_i}(x_{max}-x_{min})\Big)^2\Big] \end{aligned} \tag{2.10}$$

3. Comparison of estimators

In this section, we compare the proposed class of estimators with other existing estimators and some of their efficiency comparison conditions have been carried out under which the proposed class of estimators perform better than the other existing estimators discussed in the literature above.

(i) By (1.2) and (2.10),

$$\Big[\text{MSE}\ (\bar{y})\ -\ \text{MSE}\ (\hat{\bar{Y}}_{P_i})_{min}\Big] \geq 0, \quad \text{if}$$

$$\Big[\frac{1}{2(N-1)}\Big\{(y_{max}-y_{min}) - k_{P_i}(x_{max}-x_{min})\Big\}^2 - k_{P_i}^2 S_x^2 + 2k_{P_i}S_{yx}\Big] \geq 0.$$

(ii) By (1.4) and (2.10),

$$\Big[\text{MSE}\ (\bar{y}_s)\ -\ \text{MSE}\ (\hat{\bar{Y}}_{P_i})_{min}\Big] \geq 0, \quad \text{if}$$

$$\Big[k_{P_i}\Big(\frac{(x_{max}-x_{min})^2}{2(N-1)} - S_x^2\Big) - \Big(\frac{(y_{max}-y_{min})(x_{max}-x_{min})}{N-1} - 2S_{yx}\Big)\Big] \geq 0.$$

(iii) By (1.7) and (2.10),

$$\Big[\text{MSE}\ (\hat{\bar{Y}}_R)\ -\ \text{MSE}\ (\hat{\bar{Y}}_{P_i})_{min}\Big] \geq 0, \quad \text{if}$$

$$\Big[\frac{1}{2(N-1)}\Big((y_{max}-y_{min}) - k_{P_i}(x_{max}-x_{min})\Big)^2 + S_x^2(R-k_{P_i})(R+k_{P_i}-2\,\delta)\Big] \geq 0.$$

(iv) By (1.10) and (2.10),

$$\Big[\text{MSE}\ (\hat{\bar{Y}}_{SD})\ -\ \text{MSE}\ (\hat{\bar{Y}}_{P_i})_{min}\Big] \geq 0, \quad \text{if}$$

$$\Big[\frac{1}{2(N-1)}\Big((y_{max}-y_{min}) - k_{P_i}(x_{max}-x_{min})\Big)^2 + S_x^2(R\alpha_1-k_{P_i})(R\alpha_1+k_{P_i}-2\delta)\Big] \geq 0.$$

Where

$$\delta = \frac{\rho_{yx}S_y}{S_x}.$$

4. Numerical illustration

In this section, we illustrate the performance of the proposed class of estimators in comparison with various other existing estimators through three natural populations. The description and the necessary data statistics are given by:

Population-1: [Source: Singh and Mangat (1996), p.193]

Y: be the milk yield in kg after new food, and

X: be the yield in kg before new yield.

$$\begin{aligned} N &= 27,\ n = 12,\ \overline{X} = 10.4111,\ \overline{Y} = 11.2519,\ y_{max} = 14.8,\ y_{min} = 7.9,\ x_{max} = 14.5, \\ & x_{min} = 6.5,\ S_y^2 = 4.103\,,\ S_x^2 = 4.931,\ S_{yx} = 4.454,\ \rho_{yx} = 0.990. \end{aligned}$$

Population-2: [Source: Murthy (1967), p.399]

Y: be the area under wheat crop in 1964, and

X: be the area under wheat crop in 1963.

$$\begin{aligned} N &= 34,\ n = 12,\ \overline{X} = 208.882,\ \overline{Y} = 199.441,\ y_{max} = 634,\ y_{min} = 6,\ x_{max} = 564, \\ & x_{min} = 5,\ S_y^2 = 22564.56,\ S_x^2 = 22652.05,\ S_{yx} = 22158.05,\ \rho_{yx} = 0.980. \end{aligned}$$

Population-3: [Source: Cochran (1977), p.152]

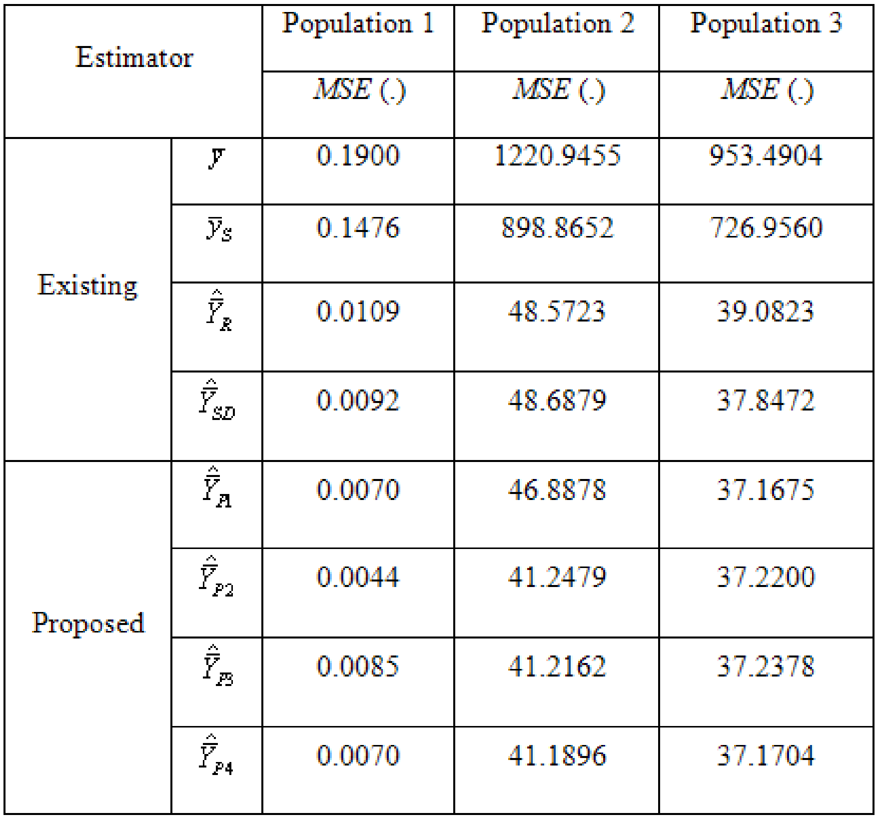
Y: be the population size in 1930 (in 1000), and

X: be the population size in 1920 (in 1000).

$$\begin{aligned} N &= 49,\ n = 12,\ \overline{X} = 103.1429,\ \overline{Y} = 127.7959,\ y_{max} = 634,\ y_{min} = 46,\ x_{max} = 507, \\ & x_{min} = 2,\ S_y^2 = 15158.83,\ S_x^2 = 10900.42,\ S_{yx} = 12619.78,\ \rho_{yx} = 0.98. \end{aligned}$$

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

Table-1: *MSE* of the Competing and the Proposed Class of Estimators



Estimator		Population 1	Population 2	Population 3
		MSE (.)	*MSE* (.)	*MSE* (.)
Existing	$\bar{y}$	0.1900	1220.9455	953.4904
	$\bar{y}_S$	0.1476	898.8652	726.9560
	$\hat{\bar{Y}}_R$	0.0109	48.5723	39.0823
	$\hat{\bar{Y}}_{SD}$	0.0092	48.6879	37.8472
Proposed	$\hat{\bar{Y}}_{P1}$	0.0070	46.8878	37.1675
	$\hat{\bar{Y}}_{P2}$	0.0044	41.2479	37.2200
	$\hat{\bar{Y}}_{P3}$	0.0085	41.2162	37.2378
	$\hat{\bar{Y}}_{P4}$	0.0070	41.1896	37.1704

For the percent relative efficiencies (PREs) of the proposed class and the existing estimators, we use the following expression for efficiency comparison. The results are, then, shown in Table-2.

$$\mathrm{PRE}\left(\hat{\bar{Y}}_g, \bar{y}\right) = \frac{\mathrm{MSE}\left(\bar{y}\right)}{\mathrm{MSE}\left(\hat{\bar{Y}}_g\right)} \times 100, \quad \text{where} \quad g = S,\ R,\ SD,\ P_1,\ P_2,\ P_3 \text{ and } P_4.$$

Table-2: *PRE* of Different Estimators with Respect to $\bar{y}$

Estimator		Population 1	Population 2	Population 3
		$PRE(.,\hat{\bar{Y}})$	$PRE(.,\hat{\bar{Y}})$	$PRE(.,\hat{\bar{Y}})$
Existing	$\bar{y}$	100.00	100.00	100.00
	$\bar{y}_S$	128.7263	135.8319	131.1621
	$\hat{\bar{Y}}_R$	1743.1193	2513.6662	2439.6988
	$\hat{\bar{Y}}_{SD}$	2065.2174	2507.6414	2519.3156
Proposed	$\hat{\bar{Y}}_{P1}$	2714.2857	2603.9727	2565.3875
	$\hat{\bar{Y}}_{P2}$	4318.1818	2960.0186	2561.7689
	$\hat{\bar{Y}}_{P3}$	2235.2941	2962.2952	2560.5444
	$\hat{\bar{Y}}_{P4}$	2714.2857	2964.2082	2565.1874

5. Conclusion

In this study, we have developed some ratio-type estimators under maximum and minimum values using knowledge of the coefficient of variation and coefficient of correlation of the auxiliary variable. We have found some theoretical possibilities under which the proposed class of estimators have smaller mean squared errors than the usual unbiased estimator; the classical ratio estimator; and the other competing estimators suggested by statisticians. Theoretical results are also verified with the help of three natural populations and their statistics are shown in table 1 and table 2, which clearly indicates that the proposed estimators have smaller mean squared errors and larger percent relative efficiency than the other estimators discussed in the literature. Thus the proposed estimators under maximum and minimum values may be preferred over the existing estimators for the use of practical applications.

6. Acknowledgements

The authors wish to thank the reviewers for their careful assessment, constructive comments and suggestions, which significantly improved this paper.

References

[1] Al-Hossain, Y. and Khan, M. *Efficiency of ratio, product and regression estimators under maximum and minimum values using two auxiliary variables*, Journal of Applied Mathematics **2014**, Atricle ID 693782, 2014.

[2] Cochran, W. G. *Sampling Techniques*, (John Wiley and Sons, New York, 1977).

[3] Das, A. K. Tripathi, T. P. *A class of sampling strategies for population mean using information on mean and variance of an auxiliary character*, Proceeding of the Indian Statistical Institute; Golden Jubilee International Conference on Statistic: Applications and New Directions, Calcutta, 16-19 December, 1981, 174-181, 1981.

[4] Das, A. K. and Tripathi, T. P. *Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known*, Sankhya-C **42**, 76-86, 1980.

[5] Kadilar, C. and Cingi, H. *A new ratio estimator in stratified random sampling*, Communication in Statistics: Theory and Methods **34**, 597-602, 2005.

[6] Kadilar, C. and Cingi, H. *An improvement in estimating the population mean by using the correlation cofficient*, Hacettepe Journal of Mathematics and Statistics **35** (1), 103-109, 2006.

[7] Khan, M. and Shabbir, J. *A ratio type estimator for the estimation of population variance using quartiles and its functions of an auxiliary variables*, Journal of Statistics Applications & Probability **2** (3), 157-162, 2013.

[8] Khan, M. and Shabbir, J. *Some improved ratio, product and regression estimators of finite population mean when using minimum and maximum values*, The Scientific World Journal, 1-7, 2013.

[9] Murthy, N. M. *Product method of estimation*, Sankhya-A **16**, 69-74, 1964.

[10] Sarndal, C. E. *Sampling survey theory vs general statistical theory: Estimation of the population mean*, International Statistical Institue **40**, 1-12, 1972.

[11] Singh, R. and Mangat, N. S. *Elements of Survey Sampling*, (Kluwer Academic Publishers, Dordrecht, Boston, London, 1996).

[12] Singh, S. *Golden and Silver Jubilee year-2003 of the linear regression estimators*, Proceeding of the American Statistical Association, Survey Method Section [CD-ROM], Toronto, Canada: American Statistical Association, 4382–4389, 2004.

[13] Sisodia, B. V. S. and Dwivedi, V. K. *A modified ratio estimator using coefficient of variation of auxiliary variable*, Journal of the Indian Society of Agricultural Statistics **33** (1), 13-18, 1981.

[14] Upadhyaya, L. N. and singh, H. P. *Use of transformed auxiliary variable in estimating the finite population mean*, Biometrical Journal **41** (5), 627–636, 1999.