Original Research Article

Variance Estimation using Linear Combination of skewness and quartiles

Abstract:- In this paper we have suggested Modified ratio type variance estimators for the estimation of population variance when there is strong correlation between auxiliary variable and study variable by using, the linear combination of skewness and quartiles as auxiliary information. To judge the efficiency of suggested estimators over existing estimators practically, we have carried out the Bias and Mean square error of proposed and existing estimators and suggested estimators have proven better performance than the existing estimators.

Comment [DA1]: 1.Re-write this sentence
2.. Include the objective of the study in the abstract

Key words: Simple random sampling, bias, mean square error, skewness and quartiles efficiency.

Introduction:

Here we consider a finite population $U = \{U_1, U_2, ..., U_N\}$ of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on $U_i, i = 1,2,3.....N$ given a vector $[Y_1, Y_2, Y_3.........Y_N]$. Sometimes in sample surveys information on auxiliary variable X correlated with study variable Y, is available can be utilized to obtain the efficient estimator for the estimation of Population variance. To estimate the population variance, various efficient estimators have been widely discussed by the authors such as Isaki [1] who proposed ratio and regression estimators. Latter various authors such as Kadilar & Cingi [2], J. Subramani & G. Kumarapandiyan [3] have also proposed estimators to improve the efficiency of the estimators over existing estimators. "In this paper our aim is to estimate the population variance $S_y^2 = \frac{1}{N-1}\sum_{i-1}^{N}(y_i - \bar{y})^2$ on the bases of random sample selected from population U" in the presence of outliers.

.MATERIALS AND METHODS:-

Notations:- $N =$ Population size. $n =$ Sample size. $\gamma = \frac{1}{n}$, Y= study variable. X= Auxiliary variable. $\bar{X}$, $\bar{Y}$ = Population means. $\bar{x}$, $\bar{y}$ = Sample means. S_Y^2 , S_x^2 = population variances. s_y^2 , s_x^2 = sample variances. C_x , C_y = . Coefficient of variation $\rho =$ coefficient $\beta_{1(x)} =$ Skewness of auxiliary variable $\beta_{2(x)} =$ Kurtosis of the auxiliary variable. $\beta_{2(y)} =$ Kurtosis of the study variable. M_d= Median of the auxiliary variable. B (.)=Bias of the estimator. MSE (.)= Mean square error. $\hat{S}_R^2 =$ Ratio type variance estimator.

$\hat{S}_{US1}^2 =$ Existing estimator proposed by Upadhyaya &Singh

$\hat{S}_{Kc1}^2$,= Existing modified ratio estimator (proposed by Kadilar & Cingi), $\hat{S}_{jG}^2$,= Existing Modified ratio estimator (proposed by J.Subramani & G.Kumarapandian), $Q_d =$ Quartile deviation, $Q_a =$ Quartile average, $Q_1 =$ first quartile, $Q_2 =$ second quartile $Q_3 =$ third quartile $Q_r = Q_3 - Q_1$ quartile range. $\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}}$

Where $\mu_{rs} = \frac{1}{N}\sum(Y_i - \bar{Y})^r (X_i - \bar{X})^s$

In this paper, we discuss the already existing estimators in the literature and then proposed modified estimators where we have used the linear combination of skewness and quartiles and accordingly we will compare the results of suggested estimators with the existing estimators

Existing Estimators:-

Ratio type Variance estimator proposed by Upadhyaya and Singh [1]

Isaki suggested a ratio type variance estimator for the population variance S_Y^2 when the population variance S_x^2 of the auxiliary variable X is known. Bias and mean square error is given by

$$\hat{S}_{US1}^2 = s_y^2 \left[\frac{S_x^2 + \beta_{2x}}{s_x^2 + \beta_{2x}} \right] \quad (1)$$

Bias $((\hat{S}_R^2) = \gamma S_y^2 \ \left[\left(\beta_{2(x)} - 1 \right) - (\lambda_{22} - 1) \ \right]$

MSE $((\hat{S}_R^2) = \gamma S_y^4 \ \left[\left(\beta_{2(y)} - 1 \right) + \ (\beta_{2(x)} - 1) - \ 2(\lambda_{22} - 1) \ \right]$

Ratio type Variance estimator proposed byKadilar and Cingi [2]

The authors have suggested ratio type variance estimators where they have used known values of Coefficient of variance and coefficient of kurtosis as an auxiliary variable X. Bias and mean square error is given by

$$\hat{S}_{kc1}^2 = s_y^2 \left[\frac{S_x^2 + C_x}{s_x^2 + C_x} \right] \quad (2)$$

Bias $((\hat{S}_{kc1}^2) = \gamma S_y^2 A_1 \ \left[A_1 \left(\beta_{2(x)} - 1 \right) - (\lambda_{22} - 1) \ \right]$

MSE $((\hat{S}_{kc1}^2) = \gamma S_y^4 \ \left[\left(\beta_{2(y)} - 1 \right) + A_1^2 \ (\beta_{2(x)} - 1) - \ 2A_1 (\lambda_{22} - 1) \ \right]$

Recent Developments:-

Comment [DA2]: Tables 1, 2 3 4,etc were not mentioned in the body of the manuscript. Please refer tables in the text

Ratio type Variance estimator proposed by J. Subramani & G. Kumara pandiyan [3]

The authors have suggested ratio type variance estimators and they have used median, quartiles and Deciles as an auxiliary variable X. Bias and mean square error is given by

$$\hat{S}_{jG}^2 = s_y^2 \left[\frac{S_x^2 + \alpha w_i}{s_x^2 \ + \alpha w_i} \right] \quad (3)$$

Bias $((\hat{S}_{jG}^2) = \gamma S_y^2 A_{jG} \ \left[A_{jG} \left(\beta_{2(x)} - 1 \right) - (\lambda_{22} - 1) \ \right]$

MSE ($(\hat{S}^2_{jG}) = \gamma S_y^4 \; \left[(\beta_{2(y)} - 1) + A^2_{jG} \; (\beta_{2(x)} - 1) - \; 2A_{jG} \, (\lambda_{22} - 1) \; \right]$

Table.1 Bias and MSE of Existing Estimators

Population-1			Population-2		Population-3		
Existing estimator	Bias	Mean square error	Bias	Mean square error	Bias	Mean square error	
Upadhyaya & Singh[1]	6.88	2678.64	189.61	8285277.52	267.48	8662043.54	
Kadilar&Cingi[2]	7.51	2806.20	193.45	8305040.92	274.87	8700066.36	
Subramani&Kumarapandiyan[3]	6.22	2551.09	194.29	8309384.71	276.27	8707285.44	

Proposed estimator:-

We have proposed a new modified ratio type variance estimator of the auxiliary variable by using linear combination of skewness and quartiles. Since quartiles are not sensitive to outliers, as they divide the series into different series and provide various location parameters and shape shifts, accounts the distributional properties and also estimates the covariate effects of average value

$$\hat{S}^2_{MS1} = s_y^2 \left[\frac{S_x^2 + (\beta_1 + Q_1)}{s_x^2 + (\beta_1 + Q_1)} \right] \qquad \hat{S}^2_{MS2} = s_y^2 \left[\frac{S_x^2 + (\beta_1 + Q_2)}{s_x^2 + (\beta_1 + Q_2)} \right] \qquad \hat{S}^2_{MS3} = s_y^2 \left[\frac{S_x^2 + (\beta_1 + Q_3)}{s_x^2 + (\beta_1 + Q_3)} \right]$$

$$\hat{S}^2_{MS4} = s_y^2 \left[\frac{S_x^2 + (\beta_1 + Q_d)}{s_x^2 + (\beta_1 + Q_d)} \right] \qquad \hat{S}^2_{MS5} = s_y^2 \left[\frac{S_x^2 + (\beta_1 + Q_a)}{s_x^2 + (\beta_1 + Q_a)} \right] \qquad \hat{S}^2_{MS6} = s_y^2 \left[\frac{S_x^2 + (\beta_1 + Q_r)}{s_x^2 + (\beta_1 + Q_r)} \right]$$

Here we have derived the bias and mean square error of the proposed estimator $\hat{S}^2_{MSi}\,;\; i = 1,2,....,6$ up to the first order of approximation as given below:

Let $e_0 = \frac{s_y^2 - S_y^2}{S_y^2}$ and $e_1 = \frac{s_x^2 - S_x^2}{S_x^2}$. Further we can write $s_y^2 = S_y^2(1 + e_0)$ and $s_x^2 = S_x^2(1 + e_0)$ and from the definition of e_0 and e_1 we obtain:

$$E[e_0] = E[e_1] = 0\,, \quad E[e_0^2] = \frac{1-f}{n}(\beta_{2(y)} - 1)\,, \quad E[e_1^2] = \frac{1-f}{n}(\beta_{2(x)} - 1)\,, \; E[e_0 e_1] = \frac{1-f}{n}(\lambda_{22} - 1)$$

The proposed estimator $\hat{S}^2_{MSi}\,;\; i = 1,2,...,6$ is given below:

$$\hat{S}^2_{MSi} = s_y^2 \left[\frac{S_x^2 + \alpha a_i}{s_x^2 + \alpha a_i} \right]$$

$\Rightarrow \quad \hat{S}^2_{MSi} = s^2_y(1+e_0)\left[\frac{S^2_x + \alpha a_i}{s^2_x + e_1 S^2_x + \alpha a_i}\right] \quad \Rightarrow \quad \hat{S}^2_{MSi} = \frac{S^2_y(1+e_0)}{(1+A_{MSi}e_1)}$ Where $A_{MSi} = \frac{S^2_x}{S^2_x + \alpha a_i}$ is a constant and MS_i, i=1,2……6 suggested estimators and $a_i = (\beta_1 + Q_i)$; $i = 1,2,3,d,a,r$

$$\Rightarrow \quad \hat{S}^2_{MSi} = S^2_y(1+e_0)(1+A_{MSi}e_1)^{-1} \quad \Rightarrow \quad \hat{S}^2_{MSi} = S^2_y(1+e_0)(1-A_{MSi}e_1 + A^2_{MSi}e^2_1 - A^3_{MSi}e^3_1 + ...)$$

Expanding and neglecting the terms more than 3rd order, we get

$$\hat{S}^2_{MSi} = S^2_y + S^2_y e_0 - S^2_y A_{MSi}e_1 - S^2_y A_{MSi}e_0 e_1 + S^2_y A^2_{MSi}e^2_1$$

$$\Rightarrow \hat{S}^2_{MSi} - S^2_y = S^2_y e_0 - S^2_y A_{MSi}e_1 - S^2_y A_{MSi}e_0 e_1 + S^2_y A^2_{MSi}e^2_1 \qquad (5)$$

By taking expectation on both sides of (5), we get

$$E(\hat{S}^2_{MSi} - S^2_y) = S^2_y E(e_0) - S^2_y A_{MSi}E(e_1) - S^2_y A_{MSi}E(e_0 e_1) + S^2_y A^2_{MSi}E(e^2_1)$$

$$Bias(\hat{S}^2_{MSi}) = S^2_y A^2_{MSi}E(e^2_1) - S^2_y A_{MSi}E(e_0 e_1)$$

$$Bias(\hat{S}^2_{MSi}) = \gamma S^2_y A_{MSi}[A_{MSi}(\beta_{2(x)} - 1) - (\lambda_{22} - 1)] \qquad (6)$$

Squaring both sides of (5) and (6), neglecting the terms more than 2nd order and taking expectation, we get

$$E(\hat{S}^2_{MSi} - S^2_y)^2 = S^4_y E(e^2_0) + S^4_y A^2_{MSi}E(e^2_1) - 2S^4_y A_{MSi}E(e_0 e_1)$$

$$MSE(\hat{S}^2_{MSi}) = \gamma S^4_y[(\beta_{2(y)} - 1) + A^2_{MSi}(\beta_{2(x)} - 1) - 2A_{MSi}(\lambda_{22} - 1)]$$

Results and Data analysis:

Numerical Illustration:

Population-1

We use the data of Murthy [13] page 228 in which fixed capital is denoted by X(auxiliary variable) and output of 80 factories are denoted by Y(study variable).we apply the proposed and existing estimators to this data set and the data statistics is given below:

$N = 80$, $n = 20$, $S_x = 8.4563$, $C_x = 0.7507$, $S_y = 18.3569$, $\bar{Y} = 51.8264$, $\beta_{1x} = 1.05$

$\beta_{2x} = 2.8664$, $\beta_{2y} = 2.2667$, $\lambda_{22} = 2.2209$, $Q_1 = 5.15$, $Q_2 = 7.5750$, $Q_3 = 16.9750$, $Q_r = 11.82$

$Q_a = 11.0625$, $Q_d = 5.9125$, $\bar{X} = 11.2624$

Population-2 We have taken population-2 from Singh & Chaudhary (1986) given in page 177

Population-2: Singh and Chaudhary [14]

$$N = 34, n = 20, \overline{Y} = 85.64, \overline{X} = 20{,}88, S_y = 73.31$$

$$S_x = 15.05, \beta_{1x} = 0.8732, \beta_{2x} = 2.91, \beta_{2y} = 13.36,$$
$$\lambda_{22} = 1.1525, Q_1 = 9.42, Q_2 = 15.0, Q_3 = 25.47,$$
$$Q_r = 16.05. Q_a = 17.45. Q_d = 16.05, C_x = 0.7205$$

Population-3 We have taken population-3 again from Singh & Chaudhary (1986) given in page 177

Population-3 Singh & Chaudhary [15]

$$N = 34, n = 20, .\overline{Y} = 85.64, \overline{X} = 19.94, S_y = 73.31, \lambda_{22} = 1.2244. S_x = 15.02, C_x = 0.7532. \beta_{2x} = 3.7257,$$
$$\beta_{2y} = 13.3666, \beta_{1x} = 1.2758, Q_1 = 9.925, Q_2 = 14.25, Q_3 = 27.8, Q_r = 17.87. Q_a = 18.86. Q_d = 8.9375$$

Table.2 Bias and MSE of Proposed Estimators

Population-1			Population-2		Population-3		
Existing estimator	Bias	Mean square error	Bias	Mean square error	Bias	Mean square error	
Upadhyaya & Singh[1]	6.88	2678.64	189.61	8285277.52	267.48	8662043.54	
Kadilar&Cingi[2]	7.51	2806.20	193.45	8305040.92	274.87	8700066.36	
Subramani&Kumarapandiyan[3]	6.22	2551.09	194.29	8309384.71	276.27	8707285.44	
MS_1	5.91	2508.57	176.24	817622.78	248.59	852774.58	
MS_2	5.32	2381.01	168.05	813275.44	238.83	847345.97	
MS_3	3.20	2125.91	153.51	805873.95	213.32	834887.79	
MS_4	4.31	2253.46	165.96	812979.38	231.92	844302.48	
MS_5	4.46	2254.46	164.49	811795.14	229.82	843064.95	
MS_6	5.75	2466.05	165.96	812929.38	250.23	853302.69	

Table.3 Percent relative efficiency of proposed estimators with existing estimators for Population-1

Estimators	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
Upadhyaya & Singh[1]	106.77	112.50	125.99	118.86	118.86	108.62
Kadilar&Cingi[2]	111.86	117.85	131.99	124.52	124.52	113.79
Subramani&Kumarapandiyan[3]	101.69	107.14	119.99	113.20	113.20	103.44

Table.4 Percent relative efficiency of proposed estimators with existing estimators for Population-2

Estimators	P1	P2	P3	P4	P5	P6
Upadhyaya & Singh[1]	1013.33	1018.75	1028.11	1019.12	1020.61	1019.18
Kadilar&Cingi[2]	1015.75	1021.18	1030.56	1021.55	1023.04	1021.61
Subramani&Kumarapandiyan[3]	1016.28	1021.71	1031.10	1022.09	1023.58	1022.15

Table.5 Percent relative efficiency of proposed estimators with existing estimators for Population-3

Estimators	P1	P2	P3	P4	P5	P6
Upadhyaya & Singh[1]	1015.74	1022.25	1037.51	1025.94	1027.44	1015.69
Kadilar&Cingi[2]	1020.20	1026.74	1042.06	1030.44	1031.95	1019.57
Subramani&Kumarapandiyan[3]	1021.05	1027.59	1042.92	1031.29	1032.81	1020.42

Conclusion:

Comment [DA3]: Please include discussion section. What are the advantages and disadvantages of these methods?

The above table reveals that our proposed estimators are more efficient than the existing estimators according to the percent relative efficiency criteria. Hence the proposed estimator may be preferred over existing estimators for use in practical applications

Literature Cited**:**

[3] Kadilar, C. & Cingi, H.(2006).Improvement in Variance estimation using auxiliary information. *Hacettepe Journal of mathematics and Statistics, 35(1).117-115.*

[4] Murthy, M N.(1967).*Sampling theory and methods. Calcutta Statistical Publishing House, India*.

[5] Subramani, J. and Kumarapandiyan, G. (2015).Generalized modified ratio type estimator for estimation of population variance. *Sri-Lankan journal of applied Statistics,vol16-1,69-90*.

[6] Kadilar and Cingi, H. 2006b. Ratio estimators for population variance in simple and stratified sampling. *Applied Mathematics and computation, 173:1047-1058*

[1] Cochran , W. G. (1977). *Sampling Techniques. Third Edition, Wiley Eastern limted.*

[2] Isaki, C.T. (1983). Variance estimation using auxiliary information.
Journal of the American Statistical Association,78,117-123.

[7] Abid, M., Abbas, N., Nazir, Z. H., 2016. Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. *Revista Colombiana de Estadistica,* **39**(1): 63-79

[8] Arcos, A., M. Rueda, M. D. Martinez, S. Gonzalez and Y. Roman. 2005. Incorporatin the auxiliary information available in variance estimation. *Applied Mathematical and Computation,* Vol **160**: 387-399.

[9] Jeelani, Iqbal and Maqbool, S. 2013.Modified ratio estimators of population mean using linear combination of coefficient of skewness and quartile deviation. *The south pacific journal of natural and applied sciences.*Vol **31**: 39-44.

[10] Kadilar, C., and Cingi,H, 2004. New ratio estimators using correlation coefficient. *Interstat,* **4**: 1-11.