Original Research Article

Variance Estimation using Linear Combination of skewness and quartiles

Abstract: In this paper we have suggested ratio type variance estimators for the estimation of population variance of study variable that has strong correlation with auxiliary variable by using skewness and quartiles as auxiliary variables to judge the efficiency of suggested estimators over existing estimators practically. The expression of Bias and Mean square error of proposed estimators have been derived up to first order approximation and the efficiency conditions have been also derived for the suggested estimators

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 Ui, i=1,2,3,…,N giving a vector $Y=\{y_1,y_2,...,y_N\}$. The goal is to estimate the population means $\bar{Y}=\frac{1}{N}\sum_{I=1}^{N}y_I$ or its variance $S_Y^2=\frac{1}{N-1}\sum_{I=1}^{N}(y_i-\bar{y})^2$ on the basis of random sample selected from a population U. Sometimes in sample surveys along with the study variable Y, information on auxiliary variable X, which is positively correlated with Y, is also available. The information on auxiliary variable X, may be utilized to obtain a more efficient estimator of the population. In this paper, our aim is to estimate the population variance on the basis of a random sample of size n selected from the population U . The problem of constructing efficient estimators for the population variance has been widely discussed by various authors such as Isaki (1983), who proposed ratio and regression estimators. Latter various authors such as Kadilar & Cingi (2006), Subramani & Kumarapandiyan (2015) improved the already existing estimators.

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. $\rho=$Correlation coefficient. $\beta_{1(x)}=$ Skewness of the 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. $\hat{S}_{Kc1}^2$, $\hat{S}_{jG}^2$, = Existing modified ratio estimators. *QD*= quartile deviation, $Q_{M.A}$ = quartile mean average, Q_1= first quartile, Q_2= second quartile, Q_3= third quartile, Q_r= quartile range.

MATERIALS AND METHODS:

In this paper, we first discuss the already existing estimators in the literature and then proposed the modified estimators using the linear combination of skewness and quartiles and results have been compared with the existing estimators

Ratio type Variance estimator proposed by Isaki (1983):

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. Its bias and mean square error are given by

$$\hat{S}_R^2 = s_y^2 \frac{S_x^2}{s_x^2} \qquad (1)$$

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

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

Kadilar and Cingi (2006) Estimators:

The authors suggested four ratio type variance estimators using known values of C.V and coefficient of kurtosis of an auxiliary variable X.

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

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

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

Recent Developments:

Subramani and Kumarapandiyan (2015) estimators where the authors used median, quartiles and Deciles of an auxiliary variable.

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

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

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

When $\alpha=0$ in equation (3), the above estimator reduces to Isaki (1983) estimator.

When $\alpha=1$ in equation (3), the above estimator reduces to Kadilar and Cingi (2006) estimator.

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 provides various location parameters and shape shifts, accounts the distributional properties and also estimates the covariate effects of average value

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

$$\hat{S}^2_{MS4} = s_y^2\left[\frac{S_x^2+(\beta_1+Q_d)}{s_x^2+(\beta_1+Q_d)}\right] \qquad \hat{S}^2_{MS5} = s_y^2\left[\frac{S_x^2+(\beta_1+Q_a)}{s_x^2+(\beta_1+Q_a)}\right] \qquad \hat{S}^2_{MS6} = s_y^2\left[\frac{S_x^2+(\beta_1+Q_r)}{s_x^2+(\beta_1+Q_r)}\right] \quad (4)$$

We have derived here the bias and mean square error of the proposed estimator $\hat{S}^2_{MSi};\ i=1,2,....,6$ to 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:

$$E[e_0]=E[e_1]=0, \quad E[e_0^2]=\frac{1-f}{n}(\beta_{2(y)}-1), \quad E[e_1^2]=\frac{1-f}{n}(\beta_{2(x)}-1), \quad E[e_0e_1]=\frac{1-f}{n}(\lambda_{22}-1)$$

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

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

$$\Rightarrow \quad \hat{S}^2_{MSi} = s_y^2(1+e_0)\left[\frac{S_x^2+\alpha a_i}{s_x^2+e_1S_x^2+\alpha a_i}\right] \quad \Rightarrow \quad \hat{S}^2_{MSi} = \frac{S_y^2(1+e_0)}{(1+A_{MSi}e_1)} \text{ where } A_{MSi} = \frac{S_x^2}{S_x^2+\alpha a_i}$$

$$a_i = (\beta_1 + Q_i);\ i = 1,2,3,d,a,r$$

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

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

$$\hat{S}^2_{MSi} = S_y^2 + S_y^2e_0 - S_y^2A_{MSi}e_1 - S_y^2A_{MSi}e_0e_1 + S_y^2A_{MSi}^2e_1^2$$

$$\Rightarrow \hat{S}^2_{MSi} - S_y^2 = S_y^2e_0 - S_y^2A_{MSi}e_1 - S_y^2A_{MSi}e_0e_1 + S_y^2A_{MSi}^2e_1^2 \qquad (5)$$

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

$$E(\hat{S}^2_{MSi} - S_y^2) = S_y^2E(e_0) - S_y^2A_{MSi}E(e_1) - S_y^2A_{MSi}E(e_0e_1) + S_y^2A_{MSi}^2E(e_1^2)$$

$$Bias(\hat{S}^2_{MSi}) = S_y^2 A_{MSi}^2 E(e_1^2) - S_y^2 A_{MSi} E(e_0 e_1)$$

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

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

$$E(\hat{S}^2_{MSi} - S_y^2)^2 = S_y^4 E(e_0^2) + S_y^4 A_{MSi}^2 E(e_1^2) - 2S_y^4 A_{MSi} E(e_0 e_1)$$

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

Results and Data analysis:

Numerical Illustration:

We use the data of Murthy (1967) 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 ,*Sx=8.4542*, *n=20*, *Cx=0.7507* , $\bar{X}$ = *11.2624*, $\beta_{2(x)}$ = *2.8664*, $\bar{Y}$ = *51.8264*, $\beta_{2(y)}$ = *2.2667*, ,
ρ = *0.9413* , $\beta_{1(x)}$ = *1.05*, λ_{22} = *2.2209*, S_y=*18.3569*, *Q1=9.318*, C_y=*0.3542*, , *G = 9.0408*, *Q2=7.5750*,
Q3= 16.975,*QD= 5.9125*, *Qa= 11.O625*, Q_R = *11.825.*

Bias and Mean Square Error of the existing and the proposed estimators

Estimators	Bias	Mean Square Error
Isaki (1983)	10.8762	3925.1622
Kadilar&Cingi(2006) 1	10.4399	3850.1552
Subramani&Kumarapandiyan(2015)	6.1235	3180.7740
Proposed (MS1)	**6.0113**	**3166.4003**
Proposed (MS2)	**6.6737**	**3258.3775**
Proposed (MS3)	**3.6235**	**2880.2490**
Proposed (MS4)	**7.3525**	**3357.1678**
Proposed (MS5)	**5.3967**	**3085.7783**
Proposed (MS6)	**5.1427**	**3053.4159**

Percent relative efficiency of proposed estimators with existing estimators

Estimators	P1	P2	P3	P4	P5	P6
Isaki (1983)	124.1598	120.4637	136.2785	116.9188	127.2016	128.5498
Kadilar&Cingi(2006) 1	121.7865	118.1617	133.6743	114.6846	124.7709	126.0933
Subramani&Kumarapandiyan(2015)	100.6134	97.6189	110.4346	94.7463	103.0791	104.1716

Conclusion:

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**:**

- Cochran , W. G. (1977). *Sampling Techniques. Third Edition, Wiley Eastern limted.*
- Isaki, C.T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association,78,117-123*.
- Kadilar, C. & Cingi, H. (2006). Improvement in Variance estimation using auxiliary information. *Hacettepe Journal of mathematics and Statistics, 35(1).117-115.*
- Murthy, M. N. (1967). *Sampling theory and methods. Calcutta Statistical Publishing House, India*.
- Sumramani, 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*.
- Wolter, K. M. (1985). *Introduction to variance estimation. Springer- Verlag.*