Improved Estimator of Finite Population Variance using Coefficient of Quartile Deviation

This study introduces a new, better, class of ratio estimators for the estimation of population variance of the study variable by using the coefficient of quartile deviation of auxiliary variable. Bias and mean square error of the proposed class of estimators are also derived. The conditions of efficiency comparison are also obtained. Simulation and different secondary data sets are used to evaluate the efficiency of proposed class of variance estimators over existing class of estimators. The empirical study shows that the suggested class of estimators is more efficient the existing class of estimators for the population variance
Keywords: Coefficient of Quartile Deviation, Natural Populations, Simple Random Sampling, Auxiliary Variable.

Introduction

Let us consider a finite population of size *N*, and *Y* be the real variable under investigation. Estimations of the unknown population parameters are used in general when the sample information is only available. The finite population variance $S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$" is based on random sample selection from population. Many forms of population variances can be found in literature. In this study, our aim is to propose and investigate a better class of estimators of population variance in simple random sampling (SRS). We consider the helping information that the auxiliary variable offers; in sampling theory, we usually get the upgraded sampling design in order to have a more accurate analysis. We consider this supplementary information to increase the accuracy of population variance; see for details Bhat et al. (2018)

The first ratio estimator for population variance was introduced by Isaki (1983) and many of the statisticians improved it in various ways for better performance. The notations used in this study are below:

N : Population size
n: Sample size
Y: Study variable
X : Auxiliary variable
$\bar{y}, \bar{x}$: Sample mean of study and auxiliary variable

---

$\bar{Y}, \bar{X}$: Population mean of study and auxiliary variable

s_y^2, s_x^2 : Sample variance

S_y^2, S_x^2 :Population variance

ρ: Coefficient of correlation

C_y, C_x :Coefficient of variations

Q_1 : The lower quartile

Q_3 : The upper quartile

Q_r :Inter-quartile range

Q_d : Semi-quartile range

Q_a : Semi-quartile average

Q_c : Coefficient of quartile deviation

$\beta_{2(y)}$: Coefficient of kurtosis of study variable

$\beta_{2(x)}$: Coefficient of kurtosis of auxiliary variable

$\hat{S}_R^2$: Traditional ratio type variance estimator

$\hat{t}_{ki}^2$: Suggested estimator

$\hat{S}_{JGi}^2$: Existing estimator

$Bias(.)$: Biases of estimators

$MSE(.)$: Mean square errors of estimators

Existing Class of Esimators

Isaki (1983) introduced a ratio (mean-per-unit) estimator of population variance when the S_x^2 population variance of supplementary variable is known. The estimator introduced by Isaki (1983) with it bias and mean squared error is given below:

$$\hat{S}_R^2 = s_y^2 \frac{s_x^2}{S_x^2} \tag{1}$$

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

$$MSE(\hat{S}_R^2) = \lambda S_y^4\left[(\beta_{2(y)} - 1) + (\beta_{2(x)} - 1) - 2(\lambda_{22} - 1)\right] \tag{3}$$

Here we have $\lambda_{22} = \dfrac{\mu_{22}}{\mu_{20}\mu_{02}}$, $\beta_{2(y)} = \dfrac{\mu_{40}}{\mu_{20}^2}$, $\beta_{2(x)} = \dfrac{\mu_{04}}{\mu_{02}^2}$ and

$\mu_{rs} = \dfrac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^r\left(x_i - \bar{X}\right)^s$.

Further, λ or γ has the same meaning, finite population correction factor. The estimation of variance plays, in general, a significant role in life sciences, as it is quite often used in sampling theory, while much effort have been made to enhance its estimated accuracy. Motivated by Kadilar and Cingi (2006), Subramani and Kumarapandiyan (2012b) suggested a class of estimators by using quartiles and some functions of quartiles of the supplementary variable, like the Inter-quartile range, Semi-quartile average and Semi-quartile range. In the following Table 1 we present some existing estimators along with their biases and means squared error (MSEs),

where $A_{JG1} = \dfrac{S_x^2}{S_x^2 + Q_1}$, $A_{JG2} = \dfrac{S_x^2}{S_x^2 + Q_3}$, $A_{JG3} = \dfrac{S_x^2}{S_x^2 + Q_r}$, $A_{JG4} = \dfrac{S_x^2}{S_x^2 + Q_d}$ and

$A_{JG5} = \dfrac{S_x^2}{S_x^2 + Q_a}$ are the constant for the estimator $\hat{S}_{JG1}^2$, $\hat{S}_{JG2}^2$, $\hat{S}_{JG3}^2$, $\hat{S}_{JG4}^2$, and $\hat{S}_{JG5}^2$ respectively.

Table 1: Existing estimators with their bias and MSE
Source (Subramani and Kumarapandiyan 2012b)

Estimators	B(.)	MSE(.)
$\hat{S}_{JG1}^2 = s_y^2\left[\dfrac{S_x^2+Q_1}{s_x^2+Q_1}\right]$	$\gamma S_y^2 A_{JG1}\left[A_{JG1}(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+A_{JG1}(\beta_{2(x)}-1)-2A_{JG1}(\lambda_{22}-1)\right]$
$\hat{S}_{JG2}^2 = s_y^2\left[\dfrac{S_x^2+Q_3}{s_x^2+Q_3}\right]$	$\gamma S_y^2 A_{JG2}\left[A_{JG2}(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+A_{JG2}(\beta_{2(x)}-1)-2A_{JG2}(\lambda_{22}-1)\right]$
$\hat{S}_{JG3}^2 = s_y^2\left[\dfrac{S_x^2+Q_r}{s_x^2+Q_r}\right]$	$\gamma S_y^2 A_{JG3}\left[A_{JG3}(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+A_{JG3}(\beta_{2(x)}-1)-2A_{JG3}(\lambda_{22}-1)\right]$
$\hat{S}_{JG4}^2 = s_y^2\left[\dfrac{S_x^2+Q_d}{s_x^2+Q_d}\right]$	$\gamma S_y^2 A_{JG4}\left[A_{JG4}(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+A_{JG4}(\beta_{2(x)}-1)-2A_{JG4}(\lambda_{22}-1)\right]$
$\hat{S}_{JG5}^2 = s_y^2\left[\dfrac{S_x^2+Q_a}{s_x^2+Q_a}\right]$	$\gamma S_y^2 A_{JG5}\left[A_{JG5}(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+A_{JG5}(\beta_{2(x)}-1)-2A_{JG5}(\lambda_{22}-1)\right]$

Proposed Classes of Estimators

In this study coefficient of quartile deviation is used for further improvement in existing estimators of population variance. The quartile deviation, which is a relative measure of dispersion, is known as the coefficient of quartile deviation. It is free of units of measurement and is a pure number (Bonett, 2006). Proposed estimators with their biases and MSEs given Table-2.

Table 2: Class of proposed estimators with their biases and MSE

Estimators	B (.)	MSE (.)
$\hat{t}_{k1}^2 = s_y^2\left[\dfrac{S_x^2Q_c^2+Q_1}{s_x^2Q_c^2+Q_1}\right]$	$\gamma S_y^2 K_1^*\left[K_1^*(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+K_1^{*2}(\beta_{2(x)}-1)-2K_1^*(\lambda_{22}-1)\right]$
$\hat{t}_{k2}^2 = s_y^2\left[\dfrac{S_x^2Q_c^2+Q_3}{s_x^2Q_c^2+Q_3}\right]$	$\gamma S_y^2 K_2^*\left[K_2^*(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+K_2^{*2}(\beta_{2(x)}-1)-2K_2^*(\lambda_{22}-1)\right]$
$\hat{t}_{k3}^2 = s_y^2\left[\dfrac{S_x^2Q_c^2+Q_r}{s_x^2Q_c^2+Q_r}\right]$	$\gamma S_y^2 K_3^*\left[K_3^*(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+K_3^{*2}(\beta_{2(x)}-1)-2K_3^*(\lambda_{22}-1)\right]$
$\hat{t}_{k4}^2 = s_y^2\left[\dfrac{S_x^2Q_c^2+Q_d}{s_x^2Q_c^2+Q_d}\right]$	$\gamma S_y^2 K_4^*\left[K_4^*(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+K_4^{*2}(\beta_{2(x)}-1)-2K_4^*(\lambda_{22}-1)\right]$
$\hat{t}_{k5}^2 = s_y^2\left[\dfrac{S_x^2Q_c^2+Q_a}{s_x^2Q_c^2+Q_a}\right]$	$\gamma S_y^2 K_5^*\left[K_5^*(\beta_{2(x)}-1)-(\lambda_{22}-1)\right]$	$\gamma S_y^4\left[(\beta_{2(y)}-1)+K_5^{*2}(\beta_{2(x)}-1)-2K_5^*(\lambda_{22}-1)\right]$

Whereas the constants are, $K_1^* = \dfrac{S_x^2Q_c^2}{S_x^2Q_c^2+Q_1}$, $K_2^* = \dfrac{S_x^2Q_c^2}{S_x^2Q_c^2+Q_3}$, $K_3^* = \dfrac{S_x^2Q_c^2}{S_x^2Q_c^2+Q_r}$

$K_4^* = \dfrac{S_x^2Q_c^2}{S_x^2Q_c^2+Q_d}$ $K_5^* = \dfrac{S_x^2Q_c^2}{S_x^2Q_c^2+Q_a}$

Efficiency of the proposed estimators

From table 1 the man square errors of existing class of estimators of population Variance can be written as:

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

where i=1, 2, 3, 4 and 5.

From Table 2, the MSEs of the proposed class of estimators for the population variance can be written as:

$$MSE(\hat{t}_{ki}^2) = \gamma s_y^4\left[\left(\beta_{2(y)} - 1\right) + K_i^{*2}(\beta_{2(x)} - 1) - 2K_i^{*2}\left(\lambda_{22} - 1\right)\right], \quad (5)$$

where i=1, 2, 3, 4 and 5.

From equations (3) and (5), the efficiency condition has been derived, according to which the proposed class of estimators shows more efficient behavior than the traditional ratio estimator for the population variance. Similarly, from equations (4) and (5), the efficiency condition is also derived, showing again that the proposed class of estimators are more efficient than the existing class of ratio estimators for the population variance as given by Subramani and Kumarapandiyan (2012b). These derived conditions are given below:

$$MSE(\hat{t}_{ki}^2) < MSE\left(\hat{S}_R^2\right) \quad \text{if} \quad \frac{\left(\beta_{2(x)} - 1\right)\left(k_i^* + 1\right)}{2} < \left(\lambda_{22} - 1\right), \quad \text{and}$$

$$MSE(\hat{t}_{ki}^2) < MSE\left(\hat{S}_{JGi}^2\right) \quad \text{if} \quad \frac{\left(\beta_{2(x)} - 1\right)\left(k_i^* + \hat{S}_{JGi}^2\right)}{2} < \left(\lambda_{22} - 1\right).$$

Empirical Study

The simulated and secondary data are used to check the efficiency of the suggested class of estimators for the population variance over the existing class of estimators. The first data set is taken from Murthy (1967), the second data set is taken from Singh and Chaudhary (1986), and the Third data set, concerning the production of rice crop for the period 1982-83 to 2014-15 in the Punjab, Pakistan, is taken from the Agricultural Statistics of Pakistan (Government of Pakistan, Ministry of Food, Agriculture and Livestock, 2014).

Table 3: Descriptives of the Secondary Data

Parameters	Population 1	Population 2	Population 3
N	80	70	33
n	20	25	10
$\bar{Y}$	51.8267	96.7000	2258.2
$\bar{X}$	11.2646	175.2671	1453.1
S_y	18.3569	60.7140	839.0
S_x	8.4563	140.8572	277.3504
ρ	0.9413	0.7293	0.9690
$\beta_{2(x)}$	2.8664	7.0952	1.7109
$\beta_{2(y)}$	2.2667	4.7596	1.5718
λ_{22}	2.2209	4.6038	1.5117
Q_1	5.1500	80.1500	1221.7

Q_3	16.9750	225.025	1714.2
A_{JG1}	0.9328	0.9960	0.9843
A_{JG2}	0.8082	0.9888	0.9782
A_{JG3}	0.8581	0.9928	0.9936
A_{JG4}	0.9236	0.9964	0.9968
A_{JG5}	0.8660	0.9924	0.9812
K_1^*	0.7986	0.9823	0.6392
K_2^*	0.6214	0.9520	0.5580
K_3^*	0.6333	0.9686	0.8146
K_4^*	0.7755	0.9840	0.8970
K_5^*	0.6486	0.9670	0.5958

For population 1, 2 and 3 of size 80, 70 and 33 respectively, the sample size, descriptive statistics and constants required to find bias and MSE of new and existing estimators (Subramani and Kumarapandiyan 2012b) of variance are calculated by using expression stated in table 1 and table 2. These values are given in Table 3. The biases and mean square errors for population 1, 2 and 3 are calculated by using these values and given in table 4 and table 5 to compare the efficiencies of the proposed and existing estimators.

Table 4: Bias of Reviewed and New Estimators

Estimators	Population 1	Population 2	Population 3
$\hat{S}^2_{JG1}$	8.1745	362.2715	13032.7778
$\hat{S}^2_{JG2}$	3.9193	353.2657	12649.3720
$\hat{S}^2_{JG3}$	5.5035	358.2204	13616.5730
$\hat{S}^2_{JG4}$	7.8272	362.7572	13818.1863
$\hat{S}^2_{JG5}$	5.7702	357.7407	12839.9972
$\hat{t}_{k1}$	3.6289	345.3350	-2577.0819
$\hat{t}_{k2}$	-0.6399	308.5025	-4516.4750
$\hat{t}_{k3}$	-0.4141	328.5025	3867.0818
$\hat{t}_{k4}$	2.9600	347.3958	8000.6724
$\hat{t}_{k5}$	-0.1135	326.5505	-3694.5464

Table 5: MSE of Reviewed and New Estimators

Estimators	Population 1	Population 2	Population 3	Simulated Results
$\hat{S}^2_{JG1}$	3480.3515	1427962.856	12548423583	3997.077
$\hat{S}^2_{JG2}$	2908.7734	1408850.951	12434849780	2629.099

$\hat{S}^2_{JG3}$	3098.2227	1419347.720	12724277603	3024.240
$\hat{S}^2_{JG4}$	3426.9869	1428997.768	12785803692	3852.322
$\hat{S}^2_{JG5}$	3133.1398	1418329.455	12491123595	3097.046
$\hat{t}_{k1}$	2878.6812	1392142.665	10311288122	2556.540
$\hat{t}_{k2}$	2668.7908	1316648.676	11003999666	2289.264
$\hat{t}_{k3}$	2662.0097	1357074.728	10399624186	2173.724
$\hat{t}_{k4}$	2813.5423	1396473.203	11199590496	2446.709
$\hat{t}_{k5}$	2657.7430	1353043.931	10623377994	2172.502

Table 4 presents the values of the biases of existing and proposed ratio type estimators. Each proposed ratio type estimator has a lower bias value compared to the bias value of the existing ratio type estimator. Similarly, each value of MSE of the new proposed estimators is also lower than the corresponding MSE value of the existing estimators, as given in Table 5

Conclusion

In this study the class of ratio (mean-per-unit) type estimators is being modified by using some other parameter of auxiliary variable. The coefficient of quartile deviation is used. The product of coefficient of quartile deviation is used with the functions of quartiles. Each estimator of the class of the proposed estimators is compared to the corresponding existing estimator. Numerical explorations were used to show the behaviour of the efficiency condition, biases and MSEs of both the existing and the new estimators. It is then concluded that the proposed estimators are more efficient.

REFERENCES

Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis.* 50 (11): 2953–2957

Isaki, C. T. (1983). Variance estimation using auxiliary information. JASA 78, 117-123.

Kadilar, C. and Cingi, H. (2006). Improvement in variance estimation using auxiliary information. *Hecet. J. Math. Stat.* 35(1): 111-115

Murthy, M. N. 1967. Sampling theory and methods, *Statistical Publishing Society*, Calcutta, India

Singh, D. and Chaudhary, F. S. (1986). Theory and analysis of sample survey designs, *New Age International Publisher*.

Subramani, J. and Kumarapandiyan, G. (2012a). Variance estimation using median of the auxiliary variable. *Int. J. Stat. Probab*. 1(3): 62-66

Subramani, J. and Kumarapandiyan, G. (2012b). Variance estimation using quartiles and their function of an auxiliary variable. *Int. J. Stat. App*. 2(5): 67-72

Subramani, J. and Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and median of an auxiliary variable. *J. Mod. App. Stat. Meth*. 12(1): 58-64

Bhat, Musaib & Maqbool, Dr Showkat & Mir, S & , Tehleel. (2018). A Robust Approach to Survey Estimation Using Non-conventional Auxiliary Population Parameters. Journal of Scientific Research and

Reports. 18. 1-6.
10.9734/JSRR/2018/40222.