International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(03), December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Modified class of estimators for estimating population variance in simple random sampling using auxiliary information

Rafia Jan,[1] Asra Jan[2] ,Tariq Jan[3]

[1,2,3]*Department of Statistics, University of Kashmir, J&K, (India)*

ABSTRACT

In this paper, an improved class of estimators for estimating population variance, using auxiliary information has been proposed. The expression for mean square error (MSE) up to the first order of approximation has been derived. From the proposed class, various estimators are derived by allocating suitable values of unknown parameters as particular members of the suggested class. The empirical study is carried out to illustrate the theoretical results.

Keywords: Exponential estimator, Auxiliary variable, Mean square error (MSE).

I.INTRODUCTION

Consider a finite population of size *N* of different units $U=\{U_1,U_2,....,U_N\}$. 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 $i=\{1,2,3,...,N\}$ and $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}y_i$ and $\bar{X}=\frac{1}{N}\sum_{i=1}^{N}x_i$ be the mean of the study as well as auxiliary variable respectively.

Also, $S_y^{\;2}=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$ and $S_x^{\;2}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{X})^2$ be the population variances of the study and the auxiliary variable respectively. Let C_y and C_x be the coefficient of variation of the study as well as auxiliary variable respectively, and ρ_{yx} be the correlation coefficient between *x* and *y*. Let $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ and $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ be the unbiased sample means of *y* and *x* respectively. Also let $\hat{s}_y^{\;2}=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$ and $\hat{s}_x^{\;2}=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ be the corresponding sample variances of the study as well as auxiliary variable respectively. Let $S_{yx}=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X})}{N-1}$, be the covariance between the respective subscripts.

Let $\lambda_{22} = \frac{\mu_{11}}{\sqrt{\mu_{02}\mu_{20}}}$ be the covariance between $S_y^{\ 2}$ and $S_x^{\ 2}$. Let $\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^{\ 2}}$ is the kurtosis for population of the study variable and $\beta_2(x) = \frac{\mu_{04}}{\mu_{02}^{\ 2}}$ be the kurtosis for population of the auxiliary variable, where $\mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^r (x_i - \bar{X})^s$ and $\gamma = \frac{1}{n}$. Also $\beta_2(y)^* = (\beta_2(y) - 1)$, $\beta_2(x)^* = (\beta_2(x) - 1)$ and $\lambda_{22}^{\ *} = (\lambda_{22} - 1)$. We ignored finite population correction (fpc) term because of ease of computation.

II. EXISTING ESTIMATORS IN LITERATURE

The usual unbiased estimator of $s_y^{\ 2}$ is

$$t_0 = s_y^{\ 2} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ and}$$

$$V(t_o) = n^{-1} S_y^{\ 4} \beta_2(y)$$

[1] proposed the ratio estimator for population variance as

$$t_1 = s_y^{\ 2} \frac{S_x^{\ 2}}{s_x^{\ 2}}$$

The mean square error (MSE) up to the first degree of approximation is given as

$$MSE(t_1) = \gamma . S_y^{\ 4}\left[\beta_2(y)^* + \beta_2(x)^* - 2\lambda_{22}^{\ *}\right]$$

[1] also presented the regression estimator for population variance using auxiliary variable as

$$T_{reg} = s_y^{\ 2} + b\left(S_x^{\ 2} - s_x^{\ 2}\right)$$

The mean square error (MSE) expression of the above estimator is given by

$$MSE(T_{reg}) = \gamma . S_y^{\ 4}\beta_2(y)^*\left[1 - \rho^2\left(s_y^{\ 2}, s_x^{\ 2}\right)\right]$$

Where, $\rho^2\left(s_y^{\ 2}, s_x^{\ 2}\right) = \frac{\left(\lambda_{22}^{\ 2}\right)^*}{\beta_2(y)^* \beta_2(x)^*}$

IJARSE
ISSN: 2319-8354

Assuming knowledge on $\beta_2(x)$, the population coefficient of kurtosis of the auxiliary variable *x*, [2] suggested the following ratio-type estimator for S_y^2 as

$$t_2 = s_y^2 \left[\frac{S_x^2 + \beta_2(x)}{s_x^2 + \beta_2(x)} \right]$$

Motivated by [3], [4], [5] and using the knowledge on $C_x, \beta_2(x)$ associated with auxiliary variable *x*, [6] envisaged the following ratio-type estimators for population variance S_y^2 as

$$t_3 = s_y^2 \left[\frac{S_x^2 - C_x}{s_x^2 - C_x} \right]$$

$$t_4 = s_y^2 \left[\frac{S_x^2 - \beta_2(x)}{s_x^2 - \beta_2(x)} \right]$$

$$t_5 = s_y^2 \left[\frac{S_x^2 \beta_2(x) - C_x}{s_x^2 \beta_2(x) - C_x} \right]$$

$$t_6 = s_y^2 \left[\frac{S_x^2 C_x - \beta_2(x)}{s_x^2 C_x - \beta_2(x)} \right]$$

Further on the lines of [7], one may define the following ratio type estimators for S_y^2 as

$$t_7 = s_y^2 \left[\frac{S_x^2 + C_x}{s_x^2 + C_x} \right]$$

$$t_8 = s_y^2 \left[\frac{S_x^2 \beta_2(x) + C_x}{s_x^2 \beta_2(x) + C_x} \right]$$

$$t_9 = s_y^2 \left[\frac{S_x^2 C_x + \beta_2(x)}{s_x^2 C_x + \beta_2(x)} \right]$$

We note that the estimators $t_i = (1,......,9)$ are members of the class of estimators of S_y^2 given by [8]

$$t_i = s_y^2 \left[\frac{aS_x^2 + b}{as_x^2 + b} \right]$$

where (*a*, *b*) are either constants or function of known parameters such as $X, C_x, \beta_1(x), \beta_2(x)$ *and* ρ_{yx} of the auxiliary variable *x*. The mean square error (MSE) up to the first order of approximation is given by

$$MSE(t_i) = \gamma . S_y^{\,4}\left[\beta_2(y)^* + \theta^2\beta_2(x)^* - 2\theta\lambda_{22}^{\;*}\right], \text{ where, } \theta = \frac{aS_x^{\,2}}{\left(aS_x^{\,2} + b\right)}$$

III. PROPOSED ESTIMATORS

Following [1] and Singh and [8], we propose the generalised exponential ratio estimator for estimating population variance given as

$$t_i^{\,*} = s_y^{\,2}\left[\frac{aS_x^{\,2} + b}{as_x^{\,2} + b}\right]^{\alpha} \qquad (1)$$

To obtain the bias and mean square error (MSE) we write

$$s_y^{\,2} = S_y^{\,2}(1+e_0),\; s_x^{\,2} = S_x^{\,2}(1+e_1)$$

Such that

$$E(e_0) = E(e_1) = 0$$

And

$$E\left(e_0^{\,2}\right) = \gamma.\beta_2(y), \quad E\left(e_1^{\,2}\right) = \gamma.\beta_2(x), \quad E(e_0e_1) = \gamma.\lambda_{22}.$$

Expressing (1) in terms of e's we have

$$t_i^{\,*} = S_y^{\,2}(1+e_0)\left\{(1+\theta e_1)^{-1}\right\}^{\alpha} \qquad (2)$$

Where

$$\theta = \frac{aS_x^{\,2}}{\left(aS_x^{\,2} + b\right)}$$

Expanding the right-hand side of (2) and neglecting the terms of e's with a power greater than two, we have

$$t_i^{\,*} = S_y^{\,2}(1+e_0)\left\{\left(1 - \alpha\theta e_1 + \frac{\alpha(\alpha+1)}{2}\theta^2 e_1^{\,2} - .....\right)\right\} \qquad (3)$$

Subtracting $S_y^{\,2}$ from (3) and taking expectations from both sides we get,

$$E\left(t_i^{\,*} - S_y^{\,2}\right) = E\left[S_y^{\,2}(e_0 - \alpha\theta e_1)\right] \qquad (4)$$

Squaring equation (4) from both sides, we get the MSE of the proposed estimator up to the first degree of approximation given as,

$$MSE\left(t_i^{\,*}\right) = \gamma . S_y^{\,4}\left[\beta_2(y)^* + \alpha\theta^2\beta_2(x)^* - 2\alpha\theta\lambda_{22}^{\;*}\right] \qquad (5)$$

To obtain the optimum value of α we differentiate the $MSE\left(t_i^*\right)$ with respect to α and equating the derivative to zero, α_{opt} thus obtained is given by,

$$\alpha_{opt} = \frac{\lambda_{22}{}^*}{\theta\beta_2(x)^*}$$

Using the value of α_{opt}, we get the $MSE_{\min}\left(t_i^*\right)$ given as

$$MSE_{\min}\left(t_i^*\right) = \gamma.S_y{}^4\beta_2(y)^*\left[1-\rho^{*2}\left(s_y{}^2, s_x{}^2\right)\right]$$

The proposed estimator t_i^* at its optimum condition is equally efficient as the regression estimator for population variance. As mentioned above, various estimators are derived by allocating suitable values of unknown parameters as particular members of the proposed class of estimators. Some members of the proposed class of estimators for different values of a and b are given in table 1.

Table 1

Some Members of Suggested estimator t^*		
Values of		Estimators
a	b	
1	1	$t_1^* = \left(s_y{}^2 \dfrac{S_x{}^2}{s_x{}^2}\right)^\alpha$
1	$\beta_2(x)$	$t_2^* = s_y{}^2\left[\dfrac{S_x{}^2+\beta_2(x)}{s_x{}^2+\beta_2(x)}\right]^\alpha$
1	$-C_x$	$t_3^* = s_y{}^2\left[\dfrac{S_x{}^2-C_x}{s_x{}^2-C_x}\right]^\alpha$

1	$-\beta_2(x)$	$t_4^* = s_y^2\left[\frac{S_x^2-\beta_2(x)}{s_x^2-\beta_2(x)}\right]^\alpha$
$\beta_2(x)$	$-C_x$	$t_5^* = s_y^2\left[\frac{S_x^2\beta_2(x)-C_x}{s_x^2\beta_2(x)-C_x}\right]^\alpha$
C_x	$-\beta_2(x)$	$t_6^* = s_y^2\left[\frac{S_x^2C_x-\beta_2(x)}{s_x^2C_x-\beta_2(x)}\right]^\alpha$
1	C_x	$t_7^* = s_y^2\left[\frac{S_x^2+C_x}{s_x^2+C_x}\right]^\alpha$
$\beta_2(x)$	C_x	$t_8^* = s_y^2\left[\frac{S_x^2\beta_2(x)+C_x}{s_x^2\beta_2(x)+C_x}\right]^\alpha$
C_x	$\beta_2(x)$	$t_9^* = s_y^2\left[\frac{S_x^2C_x+\beta_2(x)}{s_x^2C_x+\beta_2(x)}\right]^\alpha$

In addition to the above estimators, a large number of estimators can be generated from the proposed estimator t_i^* by substituting the different values of *a* and *b*. The expression of MSE of the above said estimators is given by

$$MSE_{\min}\left(t_i^*\right) = \gamma.S_y^4\beta_2(y)^*\left[1-\rho^{*2}\left(s_y^2,s_x^2\right)\right]$$

IV. EMPIRICAL STUDY

In this section, we consider the following data sets for numerical comparisons.

Population: {Source: [9]}. Let y = Level of apple production (1 unit=100 tones) and x = Number of apple trees (1 unit=100 trees).

Table 2 Data statistics

$N=104$	$n=20$	$\bar{Y}=6.254$	$\bar{X}=13931.683$	$S_y=11.670$
$S_x=23026.133$	$\beta_2(y)=16.523$	$\beta_2(x)=17.516$	$\rho=0.837$	$\lambda_{22}=14.398$
$\rho_{xy}=0.865$	$f=0.192$	$C_y=1.866$	$C_x=1.653$	$\gamma=0.05$

Table 3 shows MSE and percent relative efficiency (PRE) of the proposed estimator with respect to the existing estimators.

Table 3

Estimator	***MSE***	***PRE***	
t_0	14395.57	100	
t_1	4862.20	296.07	
t_2	4805.63	299.55	
t_3	4862.20	296.07	
t_4	4862.20	296.07	
t_5	4862.20	296.07	
t_6	4862.20	296.07	
t_7	4805.63	299.55	
t_8	4805.63	299.55	
t_9	4805.63	299.55	
$t_i^*(proposed)$	**4310.48**	**333.96**	

V. CONCLUSION

From the results of the empirical study and theoretical discussions, it is inferred that the proposed estimator t_i^*, for estimating the population variance of the study variable under the optimum condition performs better than the sample variance estimator, S_y^2 and traditional ratio type variance estimator t_i given in literature, therefore it should be preferred for the estimation of population variance.

REFERENCES

[1] Isaki CT (1983) Variance estimation using auxiliary information. J Am Stat Assoc 78:117–123.

[2] Upadhyaya LN, Singh HP (1999b) An estimator of population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. Vikram Math J 19:14–17.

[3] Sisodia BVS, Dwivedi VK (1981) A modified ratio estimator using coefficient of variation of auxiliary variable. J Indian Soc Agric Stat 33(2):13–18.

[4] Upadhyaya LN, Singh HP (1999a) Use of transformed auxiliary variable in estimating the finite population mean. Biometr J 41(5):627–636.

[5] Singh HP, Tailor, Rajesh, Tailor, Ritesh, Kakran MS (2004) An improved estimator of population mean using power transformation. J Indian Soc Agric Stat. 58(2):223–230

[6] Kadilar C, Cingi, H (2006a) Ratio estimators for the population variance in simple and stratified random sampling. Appl Math Comput 173(2):1047–1059.

[7] Kadilar C, Cingi H (2006b) Improvement in variance estimation using auxiliary information. Hacettepa J Math Stat 35(1):111–115.

[8] Singh HP, Solanki RS (2013) A new procedure for variance estimation in simple random sampling using auxiliary information. Stat Pap 54:479–497.

[9] Kadilar C, Cingi H (2007) Improvement in variance estimation in simple random sampling. Commun Stat Theory Methods 36(11):2075–2081.

[10] Cochran, W.G. (1977). *Sampling Techniques. John Wiley*.

[11] Shabbir J, Gupta S (2010) Some estimators of finite population variance of stratified sample mean. Commun Stat Theory Methods 39:3001–3008.

[13] Singh HP, Solanki RS (2011b) Efficient ratio and product estimators in stratified random sampling. Commun Stat Theory Methods (in press).