Global Journal of Mathematical Sciences: Theory and Practical.
ISSN 0974-3200 Volume 3, Number 4 (2011), pp. 369-376

http://www.irphouse.com

Efficient Estimators for Population Variance using Auxiliary Information

Subhash Kumar Yadav

Dept. of Maths & Stats, Dr. RML Avadh University,
Faizabad, U.P., India
E-mail: drskystats@gmail.com

Abstract

In this paper a family of efficient estimators, estimating population variance of the variable under study using auxiliary information has been proposed. The expressions for its bias and mean squared error (MSE) have been obtained upto $O(n^{-1})$. A comparison has been made with the general family of estimators for population variance of R. Singh et.al. (2007), which contains some well known estimators of population variance as a particular member such as Isaki (1983), Upadhyaya and Singh (1999) and Kadilar and Cingi (2006) etc. An improvement has been shown over above family of estimators through an empirical study.

Keywords: Auxiliary information, variance estimator, bias, mean squared error, efficiency.

Mathematics Subject Classification 2000: Primary 62D05 Secondary 62F10

Introduction and Notation

The use of auxiliary information may increase the precision of the estimators. When the variable under study Y is highly correlated with the auxiliary variable X, ratio and product type estimators are used for improved estimation of population parameters. Many authors have proposed ratio type estimators for the estimation of population variance using different known parameters of the auxiliary variable. Isaki (1983) was the first who used auxiliary information to estimate the variance of the variable under study. He has shown that his estimator t_1 is better than the usual estimator t_0, which does not utilize the auxiliary information in the sense of having lesser mean squared error. In the series of improvement Upadhyaya and Singh (1999) has given an estimator using the known population coefficient of kurtosis of auxiliary variable and he showed that his estimator t_6 is better than t_0 and t_1. Kadilar and Cingi (2006) proposed an estimator which utilizes the known population coefficient of variation and showed that his estimator is better than the all above estimators. In the present study, we suggest a new family of estimators for estimating population variance of the variable under study.

Material and Methods

Let the population consists of N units and a sample of size n is drawn from this population using simple random sampling without replacement. Let Y_i and X_i be the values for the ith unit ($i = 1, 2, ...,N$) of the population for the study variable and auxiliary variable respectively. Further, let $\bar{y}$ and $\bar{x}$ be the sample means of the study and auxiliary variable respectively.

In order to study the large sample properties of the proposed family of estimators, we define

$$s_y^2 = S_y^2(1+e_0) \text{ and } s_x^2 = S_x^2(1+e_1) \text{ with } E(e_i) = 0, \quad i = (0,1)$$

in case of simple random sampling without replacement, ignoring finite population correction term, the following expectations could be obtained either directly or by the method due to Kendall and Stuart (1977) as

$$E(e_0^2) = \frac{1}{n}(\lambda_{40} - 1), \; E(e_1^2) = \frac{1}{n}(\lambda_{04} - 1) \text{ and } E(e_0 e_1) = \frac{1}{n}(\lambda_{22} - 1)$$

Where $\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}}$ and $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^r (x_i - \bar{X})^s$, (r, s) = 0, 1, 2, 3, 4

Suggested family of estimators

R. Singh et.al. (2007) suggested a family of estimators for population variance as

$$t = s_y^2 \frac{(aS_x^2 - b)}{\left[\alpha(as_x^2 - b) + (1-\alpha)(aS_x^2 - b)\right]} \tag{3.1}$$

where ($a \neq 0$), b are either real numbers or the function of the known parameters of the auxiliary variable x such as coefficient of variation C_x and coefficient of kurtosis $\beta_2(x) = \lambda_{04}$.

The MSE of this family of estimators is given by

$$MSE(t) = \frac{S_y^2}{n}\left[(\lambda_{40} - 1) - 2\alpha v(\lambda_{22} - 1) + \alpha^2 v^2 (\lambda_{04} - 1)\right] \tag{3.2}$$

where $v = \dfrac{aS_x^2}{(as_x^2 - b)}$

The minimum MSE for $\alpha_{opt} = \dfrac{C}{v}$ where $C = \dfrac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$ is

$$MSE_{\min}(t) = \frac{S_y^2}{n}\left[(\lambda_{40} - 1) - (\lambda_{04} - 1)\right] \tag{3.3}$$

The ratio type estimators, given in table1 are in t-family with the MSE as

$$MSE(t_i) = \frac{S_y^2}{n}\left[(\lambda_{40} - 1) - 2\alpha v_i(\lambda_{22} - 1) + \alpha^2 v_i^2(\lambda_{04} - 1)\right] \tag{3.4}$$

$\mathrm{i} = 0, 1, \ldots, 6$

With

$v_0 = 0$, $v_1 = 1$, $v_2 = \dfrac{S_x^2}{(S_x^2 - C_x)}$, $v_3 = \dfrac{S_x^2}{(S_x^2 - \lambda_{04})}$, $v_4 = \dfrac{S_x^2\lambda_{04}}{(S_x^2\lambda_{04} - C_x)}$,

$v_5 = \dfrac{S_x^2 C_x}{(S_x^2 C_x - \lambda_{04})}$ and $v_6 = \dfrac{S_x^2}{(S_x^2 + \lambda_{04})}$.

Table 1: Some members of t-family of estimators.

Estimator	Values of		
	α	a	b
$t_0 = s_y^2$	0	0	0
$t_1 = \dfrac{s_y^2}{s_x^2} S_x^2$ Isaki (1983) estimator	1	1	0
$t_2 = \dfrac{s_y^2}{s_x^2 - C_x}[S_x^2 - C_x]$ Kadilar & Cingi (2006) estimator	1	1	C_x
$t_3 = \dfrac{s_y^2}{s_x^2 - \beta_2(x)}[S_x^2 - \beta_2(x)]$	1	1	$\beta_2(x)$
$t_4 = \dfrac{s_y^2}{\beta_2(x)s_x^2 - C_x}[\beta_2(x)S_x^2 - C_x]$	1	$\beta_2(x)$	C_x
$t_5 = \dfrac{s_y^2}{C_x s_x^2 - \beta_2(x)}[C_x S_x^2 - \beta_2(x)]$	1	C_x	$\beta_2(x)$

$t_6 = \frac{s_y^2}{s_x^2 + \beta_2(x)}[S_x^2 + \beta_2(x)]$ Upadhyaya & Singh (1999) estimator	1	1	- $\beta_2(x)$

Motivated by Nursel Koyuncu and Cem Kadilar (2009), we propose a new family of estimators for population variance as

$$\xi = ks_y^2 \frac{(aS_x^2 - b)}{[\alpha(as_x^2 - b) + (1-\alpha)(aS_x^2 - b)]} \tag{3.5}$$

where k is suitably chosen constant to be determined.

Now expressing the estimator ξ in terms of e_i $(i = 0,1)$, (2.5) can be written as

$$\xi = kS_y^2(1+e_0)(1+\alpha ve_1)^{-1} \tag{3.6}$$

Expanding the right hand side of (2.6) to the first order of approximation and subtracting S_y^2 from both the sides, we get

$$\xi = kS_y^2(1+e_0)(1-\alpha ve_1 + \alpha^2 v^2 e_1^2)$$

Now

$$\xi - S_y^2 = kS_y^2(1+e_0 - \alpha ve_1 + \alpha^2 v^2 e_1^2 - \alpha ve_0 e_1) - S_y^2 \tag{3.7}$$

Taking expectation on both sides of (2.7), we get the bias of the estimator ξ as

$$B(\xi) = \frac{kS_y^2}{n}[\alpha^2 v^2(\lambda_{04} - 1) - \alpha v(\lambda_{22} - 1)] + S_y^2(k-1) \tag{3.8}$$

Squaring both sides to equation (2.7), it gives

$$(\xi - S_y^2)^2 = k^2 S_y^4(1+e_0 - \alpha ve_1 + \alpha^2 v^2 e_1^2 - \alpha ve_0 e_1)^2 + S_y^4 - 2kS_y^4(1+e_0 - \alpha ve_1 + \alpha^2 v^2 e_1^2 - \alpha ve_0 e_1)$$

Now taking expectation both sides, we get MSE of ξ upto $O(n^{-1})$ as

$$MSE(\xi) = S_y^4[\frac{1}{n}\{k^2(\lambda_{40} - 1) + (3k^2 - 2k)\alpha^2 v^2(\lambda_{04} - 1) - 2\alpha v(2k^2 - k)(\lambda_{22} - 1)\} + (k-1)^2] \tag{3.9}$$

The minimum of $MSE(\xi)$ is obtained for optimum value of k which is $k_{opt} = \frac{A}{B}$.

Where $A = \frac{1}{n}[\alpha^2 v^2(\lambda_{04} - 1) - \alpha v(\lambda_{22} - 1)] + 1$

And $$B=\frac{1}{n}[(\lambda_{40}-1)+3\alpha^2v^2(\lambda_{04}-1)-4\alpha v(\lambda_{22}-1)]+1$$

Thus the minimum MSE of the family ξ of estimators is

$$MSE_{\min}(\xi)=S_y^4[1-\frac{A^2}{B}] \tag{3.10}$$

The ratio type estimators, given in table2 are in ξ - family with the MSE as

$$MSE(\xi_i)=S_y^4[\frac{1}{n}\{k^2(\lambda_{40}-1)+(3k^2-2k)\alpha^2v_i^2(\lambda_{04}-1)$$
$$-2\alpha v_i(2k^2-k)(\lambda_{22}-1)\}+(k-1)^2]$$

i = 1, …., 6

Table 2: Some members of ξ - family of estimators.

Estimator	Values of		
	α	A	b
$\xi_1=k\dfrac{s_y^2}{s_x^2}S_x^2$	1	1	0
$\xi_2=k\dfrac{s_y^2}{s_x^2-C_x}[S_x^2-C_x]$	1	1	C_x
$\xi_3=k\dfrac{s_y^2}{s_x^2-\beta_2(x)}[S_x^2-\beta_2(x)]$	1	1	$\beta_2(x)$
$\xi_4=k\dfrac{s_y^2}{s_x^2\beta_2(x)-C_x}[S_x^2\beta_2(x)-C_x]$	1	$\beta_2(x)$	C_x
$\xi_5=k\dfrac{s_y^2}{s_x^2C_x-\beta_2(x)}[S_x^2C_x-\beta_2(x)]$	1	C_x	$\beta_2(x)$
$\xi_6=k\dfrac{s_y^2}{s_x^2+\beta_2(x)}[S_x^2+\beta_2(x)]$	1	1	- $\beta_2(x)$

Many more estimators can be formed just by putting different values of the parameters a and b of auxiliary variable.

Efficiency comparison

To the first order of approximation, after ignoring finite population correction, the variance and MSE of estimators t_0 and t_1 respectively are given as

$$V(t_0) = \frac{S_y^4}{n}[\lambda_{40} - 1]$$
$$MSE(t_1) = \frac{S_y^4}{n}[(\lambda_{40} - 1) - 2(\lambda_{22} - 1) + (\lambda_{04} - 1)]$$

and

$$V(t_0) - MSE(t_1) = \frac{S_y^4}{n}[2(\lambda_{22} - 1) - (\lambda_{04} - 1)] > 0 \text{ if,}$$
$$2(\lambda_{22} - 1) > (\lambda_{04} - 1)$$

When this condition is attended, t_1 is more efficient than t_0.
Similarly t-family is more efficient than t_1 if,

$$MSE(t_i) < MSE(t_1), \qquad i = 2,3,4,5,6.$$

that is

$$2(\lambda_{22} - 1) < (1 + v_i)(\lambda_{04} - 1)$$

Now suggested estimators ξ_i are more efficient than t_i, (i =1,…., 6) estimators if,

$$MSE_{\min}(\xi_i) < MSE(t_i), \qquad i = 2,3,4,5,6.$$

that is

$$[1 - \frac{A^2}{B}] < \frac{1}{n}[(\lambda_{40} - 1) + v_i^2(\lambda_{04} - 1) - 2v_i(\lambda_{22} - 1)]$$

and the minimum MSE of ξ - family is less than the minimum MSE of t – family if,

$$MSE_{\min}(\xi_i) < MSE_{\min}(t_i), \qquad i = 2,3,4,5,6.$$

that is

$$[1 - \frac{A^2}{B}] < \frac{1}{n}[(\lambda_{40} - 1) - (\lambda_{22} - 1)]$$

Empirical study

We used the data in R. Singh et.al (2007), given in table3, which was early used by Kadilar and Cingi (2004) for the comparison of efficiencies of the ξ- family to t – family of estimators of R. Singh et.al (2007).

Table 3: Data Statistics.

$N = 106$,	$n = 20$,	$C_y = 4.18$,	$C_x = 2.02$,	$\bar{Y} = 15.37$,	$\bar{X} = 243.76$
$S_y = 64.25$,	$S_x = 491.89$,	$\lambda_{04} = 25.71$,	$\lambda_{40} = 80.13$,	$\lambda_{22} = 33.30$	

The MSE of ξ_i and t_i , ($i = 1,\ldots,6$) and percentage relative efficiency (PRE) each ξ_i to its corresponding t_i is given in the following table4.

Result

From the table4, we see that the proposed ξ- family of estimators is better than the t – family of estimators of R. Singh et.al (2007) in the sense of having lesser mean squared error. Table4 also shows that proposed family is much better than the estimator t_0 not utilizing auxiliary information.

Table 4: MSE and Efficiencies comparison.

t – family		ξ - family		PRE of estimator ξ_i over t_i	PRE of t_i over t_0	PRE of ξ_i over t_0
Estimator	MSE	Estimator	MSE			
t_1	$S_y^4(1.962)$	ξ_1	$S_y^4(.8252291)$	237.7522	201.6564	479.44262
t_2	$S_y^4(1.9620203)$	ξ_2	$S_y^4(.8252257)$	237.7556	201.6582	479.44459
t_3	$S_y^4(1.9619192)$	ξ_3	$S_y^4(.8251859)$	237.7548	201.6782	479.46772
t_4	$S_y^4(1.962)$	ξ_4	$S_y^4(.825229)$	237.7520	201.6565	479.44268
t_5	$S_y^4(1.9619601)$	ξ_5	$S_y^4(.8252077)$	237.7535	201.6672	479.45505
t_6	$S_y^4(1.9620805)$	ξ_6	$S_y^4(.8252723)$	237.7595	201.6347	479.41752

Conclusion

From the results of the empirical study and theoretical discussions, it is concluded that the proposed ξ- family of estimators for estimating population variance under optimum condition perform much better than the usual estimator s_y^2 and also better than t - family of estimators proposed by R. Singh et.al. (2007), which contains some well known estimators of Issaki (1983), Upadhyaya and Singh (1999) and Kadilar and Cingi (2006) etc.

References

[1] Isaki, C.T., 1983, “ Variance estimation using auxiliary information”, Jour. Amer. Stat. Assoc., 78, pp. 117-123.

[2] Kadilar, C. and Cingi, H., 2006, “Ratio estimators for the population variance in simple and stratified random sampling”, Applied Mathematics and Computation, 173, pp. 1047-1059.

[3] Kendall, M. and Stuart, A., 1977, The advanced theory of statistics, Vol I, Charles Griffin & Co., London.

[4] Koyuncu, N and Kadilar, C., 2009, “Efficient estimators for the population mean”, Hacettepe Journal of Mathematics and Statistics, 38 (2), pp. 217-225.

[5] R. Singh et.al, 2007, “Auxiliary information and a priori values in construction of improved estimators”, Renaissance High Press, U.S.A., pp. 65-73.

[6] Srivastava, S.K. and Jhajj, H.S., 1980, “A class of estimators using auxiliary information for estimating finite population variance” Sankhya, C, 42, pp. 87-96.

[7] Upadhyaya, L.N. and Singh, H.P., 1999, “An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys” Vikram Mathematical Journal, 19, pp. 14-17.