STATISTICS IN TRANSITION-new series, March 2012
Vol. 13, No. 1, pp. 37—46

MODIFIED ESTIMATORS OF POPULATION VARIANCE IN PRESENCE OF AUXILIARY INFORMATION

Dr. Rajesh Tailor[1], Balkishan Sharma[2]

ABSTRACT

This paper proposes estimator of population variance using information on known parameters of auxiliary variable. The variances of the proposed estimators are obtained. It has been shown that using modified sampling fraction the proposed estimators are more efficient than the usual unbiased estimator of population variance and usual ratio estimator for population variance under certain given conditions. Empirical study is also carried out to demonstrate the merits of the proposed estimators of population variance over other estimators considered in this paper.

Key words: Finite population variance, Bias, Mean squared error Auxiliary information and Efficiency.

1. Introduction

It is known fact that in many practical situations auxiliary information is available or may be made available in cheap cost in surveys. If this information is used intelligibly, it may give better estimators in terms of efficiency in comparison to the estimators in which auxiliary information is used.

The problem of constructing efficient estimators for the population variance S_y^2 has been widely discussed by various authors such as [3]Das and Tripathi (1978), [13]Srivastava and Jhajj (1980), [14]Upadhyay and Singh (1983), [4]Garcia and Cebrian (1996), [10]Singh S. and Joarder, A. H. (1998), [11]Singh et al. (1998), [1]Cebrian and Garcia (1997) and [12]Singh, H. P. and Singh, R. (2003). Later on [9]Singh and Tailor (2003) defined a generalized class of estimators of variance using population mean, variance, coefficient of variation of

[1] School of Studies in Statistics. Vikram University, Ujjain-456010, M.P., India. E-mail: tailorraj@gmail.com.

2 Assistant Professor in Statistics. Department of Community Medicine. Sri Aurobindo Institute of Medical Sciences, INDORE, (M. P.), India. E-mail: bksnew@rediffmail.com.

auxiliary variate and correlation coefficient between study and auxiliary variate whereas [15]Upadhyaya, L. N. and Singh, H. P. (2006) and [7]Kadilar, C. and Cingi, H. (2006) considered the problem of estimating the variance of the ratio estimator.

These motivate authors to propose modified estimators of population variance based on sampling fraction using auxiliary information.

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of size N and y be a real valued function, i. e. random variable taking the values y_i (i=1,2, . . . , N) for the i^{th} unit of the population U.

Let $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$

denote unknown population mean and population mean square of (variance) the study character y. Suppose x is an auxiliary variate which is positively correlated with study variate y taking value x_i on unit U_i. Assuming that population size N is large so that finite population terms are ignored.

The usual unbiased estimator for population variance S_y^2 is given as

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{Y})^2 \tag{1.1}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ sample mean of y,

$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ population mean of y and

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$$

When the population variance of auxiliary variate S_x^2 is known, [5]Isaki (1983) proposed a ratio estimator for population variance S_y^2 of study variate y as

$$\hat{s}_R^2 = s_y^2\left(\frac{S_x^2}{s_x^2}\right) \tag{1.2}$$

where $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$ is an unbiased estimator for population variance

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$$

The variance of s_y^2 and mean squared error of $\hat{s}_R^2$ up to the first order of approximation are given as

$$V(s_y^2) = \frac{(N-n)}{Nn} S_y^4\left[\beta_2(y) - 1\right], \qquad (1.3)$$

$$MSE(\hat{s}_R^2) = \frac{(N-n)}{Nn} S_y^4\left[\beta_2(y) + \beta_2(x) - 2h\right], \qquad (1.4)$$

where $\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}$, $\beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}$, $h = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$,

$$S_y^4 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^4 \quad \text{and} \quad \mu_{st} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^s (x_i - \bar{X})^t .$$

Here $\beta_2(y)$ and $\beta_2(x)$ are the population coefficients of kurtosis of the study variate and auxiliary variate respectively.

2. Strategy-I

Using the sampling fraction f, modified ratio type estimator for population variance of study variate y is given as

$$t_1 = f\, s_y^2 + (1-f)\, s_y^2 \frac{S_x^2}{s_x^2} \qquad (2.1)$$

To obtain the bias and mean squared error of suggested estimator t_1, we write

$s_y^2 = S_y^2(1+e_0)$ and $s_x^2 = S_x^2(1+e_1)$

such that $E(e_i) = 0$, i=0 and 1

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\beta_2(y) - 1\right) ,$$

$$E(e_1^2)=\left(\frac{1}{n}-\frac{1}{N}\right)\left(\beta_2(x)-1\right) \text{ and}$$

$$E(e_0e_1)=\left(\frac{1}{n}-\frac{1}{N}\right)\left(h-1\right) .$$

Now, the suggested estimator t_1 in terms of $e_i^{'}$s may be written as

$$t_1 = S_y^2\left[f\,(1+e_0)+(1-f)(1+e_0)(1+e_1)^{-1}\right],$$

$$E(t_1-S_y^2)=S_y^2\,E\left\{e_0-e_1(1-f)+e_1^2(1-f)-e_0e_1(1-f)\right\}, \tag{2.2}$$

$$B(t_1)=(1-f)\,S_y^2\,\left\{\beta_2(x)-h\right\}. \tag{2.3}$$

Squaring and taking expectation of both sides of equation (2.2), we get mean squared error of suggested estimator t_1, up to the first degree of approximation as

$$E(t_1-S_y^2)^2=S_y^4\,E\left\{e_0^2+(1-f)^2e_1^2-2(1-f)e_0e_1\right\}, \tag{2.4}$$

$$MSE(t_1)=\left[V(s_y^2)+(1-f)^2V(s_x^2)-2(1-f)Cov(s_y^2,s_x^2)\right], \tag{2.5}$$

where s_y^2 and s_x^2 are the population variances of the study and auxiliary variates respectively and

$$V(s_y^2)=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^4\left[\beta_2(y)-1\right],$$

$$V(s_x^2)=\left(\frac{1}{n}-\frac{1}{N}\right)S_x^4\left[\beta_2(x)-1\right] \text{ and}$$

$$Cov(s_y^2,\,s_x^2)=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2\,S_x^2\left[h-1\right].$$

3. Efficiency comparisons for t_1.

From (8.1.3), (8.1.4) and (8.2.5), it is observed that

(i) Suggested estimator t_1 would be more efficient than usual unbiased estimator for population variance s_y^2, i. e.

$MSE(t_1)$ - $V(s_y^2)$ <0 if

$$\frac{1-f}{2} < \frac{Cov(s_y^2, s_x^2)}{V(s_x^2)} \tag{3.1}$$

(ii) Suggested estimator t_1 would be more efficient than [5]Isaki (1983) ratio estimator $\hat{s}_R^2$, i. e.

$MSE(t_1)$ - $MSE(\hat{s}_R^2)$ <0 if

$$\frac{f}{2} < \left[1 - \frac{Cov(s_y^2, s_x^2)}{V(s_x^2)}\right] \tag{3.2}$$

4. Strategy-II

Another modified ratio type estimator for population variance of study variate y using the sampling fraction f, is given as

$$t_2 = \left[\left(\frac{1-f}{1+2f}\right)s_y^2 + \left(\frac{3f}{1+2f}\right)s_y^2 \frac{S_y^2}{s_x^2}\right] \tag{4.1}$$

To obtain the bias and mean squared error of suggested estimator t_2, we write $s_y^2 = S_y^2(1+e_0)$ and $s_x^2 = S_x^2(1+e_1)$

such that $E(e_i) = 0$, i=0 and 1

$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)(\beta_2(y) - 1)$, $E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)(\beta_2(x) - 1)$ and

$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N}\right)(h-1)$

Now, the suggested estimator t_2 may be written in terms of $e_i^{'}$ s as

$$t_2 = S_y^2 \left[\frac{1-f}{1+2f}(1+e_0) + \frac{3f}{1+2f}(1+e_0)(1+e_1)^{-1} \right]$$

$$E(t_2 - S_y^2) = S_y^2\, E\left\{ e_0 - e_1 \frac{3f}{(1+2f)} + e_1^2 \frac{3f}{(1+2f)} - e_0 e_1 \frac{3f}{(1+2f)} \right\} \tag{4.2}$$

$$B(t_2) = \frac{3f}{(1+2f)} S_y^2 \left\{ \beta_2(x) - h \right\} \tag{4.3}$$

Squaring and taking expectation of both sides of equation (4.2), we get Mean squared error of suggested estimator t_2, up to the first degree of approximation as

$$E(t_2 - S_y^2)^2 = S_y^4\, E\left\{ e_0^2 + \left(\frac{3f}{1+2f} \right)^2 e_1^2 - 2\left(\frac{3f}{1+2f} \right) e_0 e_1 \right\}, \tag{4.4}$$

$$MSE(t_2) = \left[V(s_y^2) + \left(\frac{3f}{1+2f} \right)^2 V(s_x^2) - 2\left(\frac{3f}{1+2f} \right) COV(s_y^2, s_x^2) \right], \tag{4.5}$$

where s_y^2 and s_x^2 are the population variances of the study and auxiliary variates respectively and expressed in previous section.

5. Efficiency comparisons for t_2

From (1.3), (1.4) and (4.5) it is observed that

(i) Suggested estimator t_2 would be more efficient than usual unbiased estimator for population variance s_y^2, i. e.

$MSE(t_2)$ - $V(s_y^2)$ <0 if

$$\frac{3f}{1+2f} < \frac{2\,Cov(s_y^2, s_x^2)}{V(s_x^2)} \tag{5.1}$$

(ii) Suggested estimator t_2 would be more efficient than Isaki (1983) ratio estimator $\hat{s}_R^2$, i. e.

$MSE(t_2)$ - $MSE(\hat{s}_R^2)$ <0 if

$$\frac{1+3f}{1+f} < \frac{2\,Cov(s_y^2, s_x^2)}{V(s_x^2)} \tag{5.2}$$

6. Strategy-III

The suggested modified ratio type estimator for population variance of study variate y is given as

$$t_3 = \left[\left(\frac{1-f}{1+3f}\right)s_y^2 + \left(\frac{4f}{1+3f}\right)s_y^2\frac{S_y^2}{s_x^2}\right] \tag{6.1}$$

To obtain the bias and mean squared error of suggested estimator t_3, we write

$s_y^2 = S_y^2(1+e_0)$ and $s_x^2 = S_x^2(1+e_1)$

such that $E(e_i) = 0$, i=0 and 1

$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)(\beta_2(y)-1)$, $E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)(\beta_2(x)-1)$ and

$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N}\right)(h-1)$

Now, the suggested estimator t_3 may be written in terms of $e_i^{'}$s as

$$t_3 = S_y^2\left[\frac{1-f}{1+3f}(1+e_0) + \frac{4f}{1+3f}(1+e_0)(1+e_1)^{-1}\right],$$

$$E(t_3 - S_y^2) = S_y^2\, E\left\{e_0 - e_1\frac{4f}{(1+3f)} + e_1^2\frac{4f}{(1+3f)} - e_0 e_1\frac{4f}{(1+3f)}\right\}, \tag{6.2}$$

$$B(t_3) = \frac{4f}{(1+3f)}S_y^2\ \{\beta_2(x) - h\}. \tag{6.3}$$

Squaring and taking expectation of both sides of equation (6.2), we get mean squared error of suggested estimator t_3, up to the first degree of approximation as

$$E(t_3 - S_y^2)^2 = S_y^4\, E\left\{ e_0^2 + \left(\frac{4f}{1+3f}\right)^2 e_1^2 - 2\left(\frac{4f}{1+3f}\right) e_0 e_1 \right\}$$

(6.4)

$$MSE(t_3) = \left[V(s_y^2) + \left(\frac{4f}{1+3f}\right)^2 V(s_x^2) - 2\left(\frac{4f}{1+3f}\right) Cov(s_y^2, s_x^2) \right] \qquad (6.5)$$

where s_y^2 and s_x^2 are the population variances of the study and auxiliary variates respectively and expressed in previous section.

7. Efficiency comparisons for t_3

From (1.3), (1.4) and (6.5) it is observed that

(i) Suggested estimator t_3 would be more efficient than usual unbiased estimator for population variance s_y^2, i. e.

$MSE(t_3)$ - $V(s_y^2)$ <0 if

$$\frac{4f}{1+3f} < \frac{2\,Cov(s_y^2, s_x^2)}{V(s_x^2)} \qquad (7.1)$$

(ii) Suggested estimator t_3 would be more efficient than [5]Isaki (1983) ratio estimator $\hat{s}_R^2$, i. e.

$MSE(t_3)$ - $MSE(\hat{s}_R^2)$ <0 if

$$\frac{1+4f}{1+2f} < \frac{2\,Cov(s_y^2, s_x^2)}{V(s_x^2)} \qquad (7.2)$$

Section 3, 5 and 7 provided the conditions under which proposed estimators t_1, t_2 and t_3 have less mean squared errors in comparison to usual unbiased estimator for population variance and ratio estimator for population variance.

8. Empirical study

To analyze the performance of the proposed estimators t_1, t_2 and t_3 in comparison to other estimators, we consider the data given in [8]Murthy (1967, p.-226). The variates and data set is given as

y : Output and

x : number of workers.

N=80, n=30,

$\beta_2(y)$ =2.2667, $\beta_2(x)$ =3.65 and h=2.3377.

Table 8.1. Percent Relative Efficiencies of $\hat{s}_R^2$, t_1, t_2 and t_3 with respect to s_y^2.

PRE's	Estimators				
	s_y^2	$\hat{s}_R^2$	t_1	t_2	t_3
Percent Relative Efficiencies	100	102.05	150.53	155.75	146.88

It is observed from the table 8.1 that there is a significant gain in efficiency by using proposed variance estimators t_1, t_2 and t_3 in comparison to unbiased estimator for population variance s_y^2 and ratio estimator for population variance $\hat{s}_R^2$ given by [5]Isaki (1983).

Therefore suggested estimators t_1, t_2 and t_3 are recommended for their use in practice.

REFERENCES

CEBRIAN, A.A. and GARCIA, R.M. (1997). Variance estimation using auxiliary information. An almost unbiased multivariable ratio estimator. Metrika, 45, 171-178.

COCHRAN, W.G. (1977). Sampling Techniques. Third U. S. Edition. Wiley Eastern Limited, 325.

DAS, A.K. and TRIPATHI T.P. (1978). Use of auxiliary information in estimating the finite population variance. Sankhya, C, 40, 139-148.

GARCIA, M. RUEDA and CEBRIAN, A.A. (1996). Repeated substitution method. The ratio estimator of the population variance. Metrika, 43, 101-105.

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

KADILAR, C. and CINGI, H. (2006). Ratio estimators for population variance in simple and stratified sampling, Applied Mathematics and Computation 173, 1047–1058, 2006.

KADILAR, C. and CINGI, H. (2006). Improvement in variance estimation using auxiliary information, Hacettepe Journal of Mathematics and Statistics 35 (1), 111–115.

MURTHY, M.N. (1967). Sampling theory and methods, statistical publishing society, Calcutta.

SINGH, H. P. and TAILOR, R. (2003). Use of known correlation coefficient in estimating the finite population mean, "Statistics in Transition", 6, 555-560.

SINGH, S. and JOARDER, A.H. (1998). Estimation of finite population variance using random non-response in surveys sampling. Metrika,47, 241-249.

SINGH, H.P. UPADHYAYA, L.N. NAMJOSHI, U.D. (1988). Estimation of finite population variance, Current Science 57, 1331–1334.

SINGH, H.P. and SINGH, R. (2003). Estimation of variance through regression approach in two phase sampling. Alig. Jour. Stat., 23, 13-30.

SRIVASTAVA, S.K. AND JHAJJ H.S. (1980). A class of estimators using auxiliary information for estimating finite population variance. Sankhya , C, 42, 87-96.

UPADHYAYA, L.N. AND SINGH, H.P. (1983). Use of auxiliary information in the estimation of population variance. Mathematical forum, 4, 2, 33-36.

UPADHYAYA, L.N. and SINGH, H.P. (2006). Almost unbiased ratio and product-type estimators of finite population variance in sample surveys, Statist. in Transi.,7 5, 1087–1096.