Some Classes of Log Type Estimators Using Auxiliary Attribute for Population Variance

Shashi Bhushan[a] and Chandni Kumari[b]

[a]Department of Mathematics & Statistics, Dr. Shakuntala Mishra National Rehabilitaion University, Lucknow, 226017, India.
[b]Department of Applied Statistics, Bahasaheb Bhimrao Ambedkar University, Lucknow, 226025, India.

Abstract

This paper introduces some new classes of log type estimators assuming that the information on an auxiliary attribute related with the study variable is given. The mean squared error of the proposed classes of estimators has been obtained up to the first order of approximation. An empirical study is given as an illustration.

Keywords: Simple random sampling without replacement, Ratio method of estimation, Mean squared error, Efficiency.

1 Introduction

We use auxiliary information to improve the efficiency of the estimators. In this paper, we make use of auxiliary attribute to enhance the precision of proposed logarithmic ratio type estimator. Various authors like [20], [18], [16], [15], [12], [13], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] gave contribution in this field but the logarithmic estimator is used probably for the first time for the estimation of population variance.

Consider a finite population $U = U_1, U_2, ..., U_N$ of size N from which a sample of size n is drawn according to simple random sampling without replacement. Let, y_i and f_i denotes the value of the study and auxiliary attribute for the ith unit $i = 1, 2, ..., N$ of the population. Further, let $\bar{y}$ and p be the sample means and $s_y^2 = n^{-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ and $s_f^2 = n^{-1}\sum_{i=1}^{n}\left(f_i - p\right)^2$ be the sample variance of the study variable and auxiliary attribute respectively.

2 The suggested generalized class of log-type estimators

We propose the following new classes of log type estimators for the population variance S_y^2 as

$$T_1 = s_y^2\left[1+\log\left(\frac{S_f^2}{s_f^2}\right)\right]^{a_1} \tag{2.1}$$

$$T_2 = s_y^2\left[1+a_2\log\left(\frac{S_f^2}{s_f^2}\right)\right] \tag{2.2}$$

$$T_3 = s_y^2\left[1+\log\left(\frac{S_f^{2*}}{s_f^{2*}}\right)\right]^{a_3} \tag{2.3}$$

$$T_4 = s_y^2\left[1+a_4\log\left(\frac{S_f^{2*}}{s_f^{2*}}\right)\right] \tag{2.4}$$

where $S_f^{2*} = aS_f^2 + b$ and $s_f^{2*} = as_f^2 + b$

such that $a(\neq 0)$, b are either real numbers or functions of the known parameters of the auxiliary attribute f such as the standard deviations S_f, coefficient of variation C_f, coefficient of kurtosis b_{2f}, coefficient of skewness b_{1f} and correlation coefficient ρ of the population.

Theorem 1 *The bias of the proposed estimators are*

$$Bias(T_1) = S_y^2\ I\left[\frac{a_1^2}{2}b_{2f}^* - a_1 I_{22}^*\right]$$

$$Bias(T_2) = S_y^2\ I\left[\frac{a_2}{2}b_{2f}^* - a_2 I_{22}^*\right]$$

$$Bias(T_3) = S_y^2\ I\left[\frac{a_3^2}{2}\eta^2 b_{2f}^* - a_3\eta I_{22}^*\right]$$

$$Bias(T_4) = S_y^2\ I\left[\frac{a_4}{2}\eta b_{2f}^* - a_4\eta I_{22}^*\right]$$

where $\eta = \dfrac{a S_f^2}{a S_f^2 + b}$

Proof. Consider the estimator

$$T_1 = s_y^2\left[1+\log\left(\frac{S_f^2}{s_f^2}\right)\right]^{a_1}$$

$$= S_y^2\left(1+\varepsilon_0\right)\left[1+\log\left(1+\varepsilon_1\right)^{-1}\right]^{a_1}$$; where $\varepsilon_0 = \left(s_y{}^2 - S_y{}^2\right)/S_y{}^2$ and $\varepsilon_1 = \left(s_f^2 - S_f^2\right)/S_f^2$

$$T_1 - S_y^2 = S_y^2\left[\varepsilon_0 - a_1\varepsilon_0\varepsilon_1 - a_1\varepsilon_1 + \frac{a_1^2\,\varepsilon_1^2}{2}\right] \qquad (2.5)$$

Taking expectation on both the sides, we get

$$E\left(T_1 - S_y^2\right) = S_y^2\left[a_1\,E\left(\varepsilon_0\varepsilon_1\right) + E\left(\frac{a_1^2\,\varepsilon_1^2}{2}\right)\right]$$

Using the results of simple random sampling, we have

$E\left(\varepsilon_0{}^2\right) = Ib_{2y}^*$, $E\left(\varepsilon_1{}^2\right) = Ib_{2f}^*$ and $E\left(\varepsilon_0\varepsilon_1\right) = I\,I_{22}^*$

$$Bias\left(T_1\right) = S_y^2\ I\left[\frac{a_1^2}{2}b_{2f}^* - a_1\,I_{22}^*\right] \qquad (2.6)$$

Proceeding in a similar way, we have proved the bias of remaining proposed estimators.

Theorem 2 *The mean squared error of the proposed estimator considered up to the terms of order n^{-1} are given as follows*

$$MSE\left(T_i\right) = S_y^4\ I\left\{b_{2y}^* + a_i^2 b_{2f}^* - 2a_i\,I_{22}^*\right\} \qquad \text{for } i = 1, 2$$

$$MSE\left(T_j\right) = S_y^4\ I\left\{b_{2y}^* + a_j^2\,\eta^2 b_{2f}^* - 2a_j\,\eta I_{22}^*\right\} \qquad \text{for } j = 3, 4$$

Proof. Consider the estimator,

$$T_1 = s_y^2\left[1+\log\left(\frac{S_f^2}{s_f^2}\right)\right]^{a_1}$$

$$T_1 - S_y^2 = S_y^2\left[\varepsilon_0 - a_1\varepsilon_1\right]$$

Squaring on both the sides and taking expectation, we get

$$E\left(T_1 - S_y^2\right)^2 = S_y^4\left[E\left(\varepsilon_0{}^2\right) + a_1{}^2 E\left(\varepsilon_1{}^2\right) - 2a_1 E\left(\varepsilon_0\varepsilon_1\right)\right]$$

$$MSE\left(T_1\right) = S_y^4\ I\left\{b_{2y}^* + a_1^2 b_{2f}^* - 2a_1\,I_{22}^*\right\} \qquad (2.7)$$

On differentiating the above expression with respect to a_1, we get the optimum value of a_1. It is given by

$$a_{1opt} = I^*_{22} \big/ b^*_{2f} \qquad (2.8)$$

The optimum mean squared error is as follows

$$MSE\left(T_1\right)_{opt} = S_y^4 \; I\left\{b^*_{2y} - \frac{I^{*2}_{22}}{b^*_{2f}}\right\}$$

Proceeding in a similar way, we can obtain the mean squared error of the remaining estimators.

3 Some members of the class of estimators proposed estimators

It can be easily seen that the classes T_3 and T_4 are more generalized form of class of estimators in the sense of the constants $a(\neq)0$, b which are either real numbers or functions of the known parameters of the auxiliary attribute f such as the standard deviations S_f, coefficient of variation C_f, coefficient of kurtosis b_{2f}, coefficient of skewness b_{1f} and correlation coefficient ρ of the population. Therefore, a wide variety of estimators can be designed using the above known population parameter. Some of them are given below.

Table 1: Some generalized members of the proposed class of estimators

Log type estimators		**a**	**b**
T_3	T_4		
$T_{31} = s_y^2\left[1+\log\left(\frac{S_f^2}{s_f^2}\right)\right]^{a_3}$	$T_{41} = s_y^2\left[1+a_4\log\left(\frac{S_f^2}{s_f^2}\right)\right]$	1	0
$T_{32} = s_y^2\left[1+\log\left(\frac{S_f^2+C_f}{s_f^2+C_f}\right)\right]^{a_3}$	$T_{42} = s_y^2\left[1+a_4\log\left(\frac{S_f^2+C_f}{s_f^2+C_f}\right)\right]$	1	C_f
$T_{33} = s_y^2\left[1+\log\left(\frac{b_{2f}S_f^2+C_f}{b_{2f}s_f^2+C_f}\right)\right]^{a_3}$	$T_{43} = s_y^2\left[1+a_4\log\left(\frac{b_{2f}S_f^2+C_f}{b_{2f}s_f^2+C_f}\right)\right]$	b_{2f}	C_f
$T_{34} = s_y^2\left[1+\log\left(\frac{C_f\,S_f^2+b_{2f}}{C_f\,s_f^2+b_{2f}}\right)\right]^{a_3}$	$T_{44} = s_y^2\left[1+a_4\log\left(\frac{C_f\,S_f^2+b_{2f}}{C_f\,s_f^2+b_{2f}}\right)\right]$	C_f	b_{2f}
$T_{35} = s_y^2\left[1+\log\left(\frac{S_f^2+S_f}{s_f^2+S_f}\right)\right]^{a_3}$	$T_{45} = s_y^2\left[1+a_4\log\left(\frac{S_f^2+S_f}{s_f^2+S_f}\right)\right]$	1	S_f

$T_{36} = s_y^2 \left[1 + \log\left(\frac{b_{1f} S_f^2 + S_f}{b_{1f} s_f^2 + S_f}\right)\right]^{a_3}$	$T_{46} = s_y^2 \left[1 + a_4 \log\left(\frac{b_{1f} S_f^2 + S_f}{b_{1f} s_f^2 + S_f}\right)\right]$	b_{1f}	S_f
$T_{37} = s_y^2 \left[1 + \log\left(\frac{b_{2f} S_f^2 + S_f}{b_{2f} s_f^2 + S_f}\right)\right]^{a_3}$	$T_{47} = s_y^2 \left[1 + a_4 \log\left(\frac{b_{2f} S_f^2 + S_f}{b_{2f} s_f^2 + S_f}\right)\right]$	b_{2f}	S_f
$T_{38} = s_y^2 \left[1 + \log\left(\frac{s_f^2 + \rho}{s_f^2 + \rho}\right)\right]^{a_3}$	$T_{48} = s_y^2 \left[1 + a_4 \log\left(\frac{S_f^2 + \rho}{s_f^2 + \rho}\right)\right]$	1	ρ
$T_{39} = s_y^2 \left[1 + \log\left(\frac{S_f^2 + b_{2f}}{s_f^2 + b_{2f}}\right)\right]^{a_3}$	$T_{48} = s_y^2 \left[1 + a_4 \log\left(\frac{S_f^2 + b_{2f}}{s_f^2 + b_{2f}}\right)\right]$	1	b_{2f}
$T_{3_{10}} = s_y^2 \left[1 + \log\left(\frac{C_f S_f^2 + \rho}{C_f s_f^2 + \rho}\right)\right]^{a_3}$	$T_{4_{13}} = s_y^2 \left[1 + a_4 \log\left(\frac{C_f S_f^2 + \rho}{C_f s_f^2 + \rho}\right)\right]$	C_f	ρ
$T_{3_{11}} = s_y^2 \left[1 + \log\left(\frac{\rho S_f^2 + C_f}{\rho s_f^2 + C_f}\right)\right]^{a_3}$	$T_{4_{13}} = s_y^2 \left[1 + a_4 \log\left(\frac{\rho S_f^2 + C_f}{\rho s_f^2 + C_f}\right)\right]$	ρ	C_f
$T_{3_{12}} = s_y^2 \left[1 + \log\left(\frac{b_{2f} S_f^2 + \rho}{b_{2f} s_f^2 + \rho}\right)\right]^{a_3}$	$T_{4_{12}} = s_y^2 \left[1 + a_4 \log\left(\frac{b_{2f} S_f^2 + \rho}{b_{2f} s_f^2 + \rho}\right)\right]$	b_{2f}	ρ
$T_{3_{13}} = s_y^2 \left[1 + \log\left(\frac{\rho S_f^2 + b_{2f}}{\rho s_f^2 + b_{2f}}\right)\right]^{a_3}$	$T_{4_{13}} = s_y^2 \left[1 + a_4 \log\left(\frac{\rho S_f^2 + b_{2f}}{\rho s_f^2 + b_{2f}}\right)\right]$	ρ	b_{2f}

4 Comparison with available estimator

Let us now compare the proposed classes of estimators with the conventional estimators.

4.1 Ratio estimator using auxiliary attribute

$$t_0 = s_y^2 \left(\frac{S_f^2}{s_f^2}\right)$$

Its mean squared error is given by

$$MSE(t_0) = I\, S_y^4 \left\{b_{2y}^* + b_{2f}^* - 2\, I_{22}^*\right\} > MSE(T_1)_{opt}$$

4.2 Bahl and Tuteja type ratio estimator using auxiliary attribute

$$t_1 = s_y^2 \exp\left(\frac{S_f^2 - s_f^2}{S_f^2 + s_f^2}\right)$$

Its mean squared error is given by

$$MSE(t_1) = I\, S_y^4 \left\{ b_{2y}^* + \frac{b_{2f}^*}{4} - 2\, I_{22}^* \right\} > MSE(T_1)_{opt}$$

4.3 Das and Tripathi type ratio estimator using auxiliary attribute

$$t_2 = s_y^2 \left(\frac{S_f^2}{S_f^2 + \beta\left(s_f^2 - S_f^2\right)} \right)$$

Its mean squared error is given by

$$MSE\left(t_2\right)_{opt} = I\, S_y^4 \left\{ b_{2y}^* - \frac{I_{22}^{*2}}{b_{2f}^*} \right\} = MSE\left(T_1\right)_{opt}$$

4.4 Linear regression estimator

$$t_3 = s_y^2 + d\left(s_f^2 - S_f^2\right)$$

where d is the sample regression constant. The mean squared error is given by

$$MSE\left(t_3\right)_{opt} = I\, S_y^4 \left\{ b_{2y}^* - \frac{I_{22}^{*2}}{b_{2x}^*} \right\} = MSE\left(T_1\right)_{opt}$$

4.5 Prasad and Singh estimator

Prasad and Singh (1992) introduced the following estimator

$$t_4 = a\, s_y^2 \left(\frac{S_f^2}{s_f^2} \right)$$

where a is the sample regression constant. The mean squared error is given by

$$MSE\left(t_4\right)_{opt} = MSE\left(T_1\right)_{opt}$$

4.5 Garcia and Cebrian estimator

Garcia and Cebrian (1996) introduced the following estimator

$$t_5 = s_y^2 \left(\frac{S_f^2}{s_f^2} \right)^b$$

where b is the sample regression constant. The mean squared error is given by

$$MSE\left(t_5\right)_{opt} = MSE\left(T_1\right)_{opt}$$

4.6 Upadhaya and Singh estimator

ISSN 2229-5518

Upadhaya and Singh (2001) suggested following estimator

$$t_6 = s_y^2 + \alpha\left(S_f^2 - s_f^2\right)$$

where α is the sample regression constant. The mean squared error is given by

$$MSE\left(t_6\right)_{opt} = MSE\left(T_1\right)_{opt}$$

4.7 Yadav and Kadilar (2013) estimator

Yadav and Kadilar (2013) introduced the following estimator

$$t_7 = s_y^2 \exp\left(1 - \frac{\gamma S_f^2}{S_f^2 + \left(\gamma - 1\right) s_f^2}\right)$$

where γ is the sample regression constant. The mean squared error is given by

$$MSE\left(t_7\right)_{opt} = MSE\left(T_1\right)_{opt}$$

5 Empirical study

To compare the efficiency of the suggested class of estimator numerically, we considered nine natural data sets. The description of the population is given below.

Population 1. (Singh D and Chaudhary F. S., Pg. no. 141).

y : number of bearing lime trees

f : area under lime (in acres)

S_y^2 = 6564586.45, S_f^2 =1092.1024, b_{2y}^* = 12.2574, b_{2f}^* =4.5788, I_{22}^* = 6.7126, C_f =1.4273, P = 22.6209, ρ =0.9021, n =9

Population 2. (Choudhary F. S. and Singh D., Pg. no. 176).

y : number of cows in milk enumerated

f : proportion of cows in milk in the previous year.

S_y^2 = 332721.2079, S_f^2 = 281472.7868, b_{2y}^* = 6.2079, b_{2f}^* = 500.43, I_{22}^* = 4.9528, C_f = 0.8276, P = 641.05, ρ =0.8933, n =8

Population 3. (Singh S., Pg. no. 324-325)

y : approximate duration of sleep (in minutes)

f : age in years of the persons.

S_y^2 = 3582.579, S_f^2 = 85.2367, b_{2y}^* = 1.6678, b_{2f}^* = 1.2389, I_{22}^* = 0.9961, C_f = 0.139, P = 67.2667, ρ =0.8552, n =9

By using the above data set, the percent relative efficiency of the different estimator are given in Table 2.

Table 2: PRE of the estimators

Estimator	Pop 1	Pop 2	Pop 3
t_0	100	100	100
t_1	111.506	112.114	324.801
t_2	76.207	72.252	243.531
t_3	111.506	112.114	324.801
t_4	111.506	112.114	324.801
t_5	111.506	112.114	324.801
t_6	111.506	112.114	324.801
t_7	111.506	112.114	324.801
$T_{1_{opt}}$	111.506	112.114	324.801

From the above table, it is clear that the proposed estimator is equally efficient as compared to other estimators for all the data sets given here.

6 Conclusion

The purpose of current work is to develop some log type estimators utilizing auxiliary information in the form of attribute. The proposed estimators are found better than various ratio and product type estimators. The numerical study shows the appropriateness of proposed estimators over regression estimator, Singh et al. (1973), Das & Tripathi (1978), Prasad & Singh (1992), Garcia & Cebrian (1996), Upadhaya & Singh (2001), Yadav & Kadilar (2013) estimators.

References

[1] Bhushan S. and Kumari C. (2018). A new log type estimators for estimating the population variance, Int. J. Comp. App. Math., 13 (1), 43-54.

[2] Bhushan S., Misra P. K. and Yadav S. K. (2017a). On The Class of Double Sampling Exponential Ratio Type Estimator Using Auxiliary Information on an Attribute and an Auxiliary Variable, Int. J. Comp. App. Math., 12(1), 1-10.

[3] Bhushan S., Misra P. K. and Yadav S. K. (2017b). A Generalized Class of Unbiased Estimators for Population Mean Using Auxiliary Information on an Attribute and an Auxiliary Variable, Int. J. Comp. App. Math., 12(1), 11-28.

[4] Bhushan S., Misra P. K. and Yadav S. K. (2017c). On The Class of Double Sampling Ratio Estimators Using Auxiliary Information on an Attribute and an Auxiliary Variable, Int. J. Stat. Syst., 12(1), 15-23.

[5] Bhushan S., Misra P. K. and Yadav S. K. (2017d). A Generalized Class of Double Sampling Estimators Using Auxiliary Information on an Attribute and an Auxiliary Variable, Int. J. Comp. Intel. Res., 13(1), 23-33.

[6] Bhushan S., Misra, P. K. and Yadav S. K. (2017e). On Unbiased Class of Ratio Estimator for Population Mean Using Auxiliary Information on an Attribute and a Variable, Int. J. Stat. Syst., 12(1), 25-32.

[7] Bhushan S., Misra P. K. and Yadav S. K. (2017f). An Unbiased Estimator For Population Mean Using An Attribute And An Auxiliary Variable, Adv. Dy. Sys. Appl., 12(1), 29-39.

[8] Bhushan S. and Gupta R. (2015). An unbiased class of log-type estimators for population mean using auxiliary information on an attribute and a variable, Math. Sci. Int. Res. J., 4(1), 143-147.

[9] Bhushan S., Gupta R. and Pandey S. K. (2015). Some log-type classes of estimators using auxiliary information, Int. J. Ag. Stat. Sci., 11(2), 487 - 491.

[10] Bhushan S. (2013). Improved sampling strategies in finite population, Sch. Press, Germany.

[11] Gupta S. and Shabbir J. (2008). Variance estimation in simple random sampling using auxiliary information, Hacet. J. Math. Stat., 37, 57-67.

[12] Kadilar C. and Cingi H. (2006a). Improvement in variance estimation using auxiliary information, Hacet. J. Math. Stat., 1(35), 111-115.

[13] Kadilar C. and Cingi H. (2006b). Ratio estimators for population variance in simple and stratied sampling, App. Math. Comp., 1(73), 1047-1058.

[14] Hidiroglou M. A. and Sarndal C. E. (1998). Use of auxiliary information for two-phase sampling , Surv. Methodol., 24(1), 11-20.

ISSN 2229-5518
[15] Swain A. K. P. C. and Mishra G. (1994). Estimation of population variance under unequal probability sampling , Sankhya, B (56), 374-384.

[16] Bahl S. and Tuteja R. K. (1991). Ratio and Product type exponential estimator, Info. Optim. Sci., XII(I), 159-163.

[17] Sukhatme P. V., Sukhatme B. V., Sukhatme S. and Ashok C. (1984). Sampling Theory of Surveys with Applications , Iowa State Univ. Press, Ames.

[18] Isaki C. T. (1983). Variance estimation using Auxiliary Information , J. Amer. Stat. Asso., 78, 117-123.

[19] Chaudhury A. (1978). On estimating the variance of a finite population. Metrika, 25, 66-67.

[20] Das A. K. and Tripathi T. P. (1978). Use of auxiliary information in estimating the nite population variance. Sankhya, C(4), 139-148.

[21] Cochran W. G. (1963). Sampling Techniques , Wiley Eastern Private Limited, New Delhi.

[22] Neyman J. (1938). Contribution to the theory of sampling human populations, J. Amer. Stat. Asso., 33, 101-116.