ISSN 2229-5518

Estimation of Variance of Finite Population Using Double Sampling Scheme

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

In this paper, a class of double sampling log type estimators are proposed for estimating the population variance. The expressions for bias and mean squared error are obtained up to the first order of approximation. It has been found that the proposed estimator is most efficient than the estimators available in literature. A numerical study is also given in support of the work done.

1 Introduction

Double sampling is used when the information about auxiliary variable is not readily available. Here, we use two auxiliary information for estimating the population variance. Both the auxiliary variables are correlated with the study variable. The ratio, regression, product and difference methods take advantage of the auxiliary information at the estimation stage. Many authors like, Pandey and Dubey (1988), Upadhyaya and Singh (1999) , Kadilar and Cingi (2003), Singh and Taylor (2003), Singh (2003), Sisodia and Dwivedi (1981), Koyuncu and Kadilar along with many others have proposed various estimators using auxiliary information on various population parameters like coefficient of skewness, kurtosis, variation, standard deviation, correlation coefficient, etc. Sometimes, it is more economical to obtain information on more than one auxiliary information; this would probably help in improvising the efficiency of the estimator, used to estimate the parameter under consideration. The literature deals with a wide range of ratio, product, difference and exponential estimators proposed by various renowned authors using multiple auxiliary information (Olkin (1958), Raj (1965), Singh (1967),Shukla(1966), etc.). Recently, Bhushan and Kumari (2018) had made the use of logarithmic relationship between the study variable and auxiliary variable, we have made the use of multiple auxiliary variables x0s for estimating the population variance. The proposed estimators would work in case when the study variable is logarithmically related to the auxiliary variable.

ISSN 2229-5518

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 (SRSWOR). Let y_i, x_{i_1} and x_{i_2} denotes the value of the study and two auxiliary variable for the ith unit $i = 1, 2, ..., N$ of the population. Further, let $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$ be the sample means of study variable and two auxiliary variables. Also, $s_y^2 = N^{-1}\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2$, $s_{x_1}^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_1\right)^2$ and $s_{x_2}^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_2\right)^2$ be the sample variance of the study and two auxiliary variables respectively.

2 The suggested generalized class of log-type double sampling estimators

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

$$T_1 = w_1 s_y^2 \left[1 + \log\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\right]^{a_1} \left[1 + \log\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)\right]^{a_2} \tag{2.1}$$

$$T_2 = w_2 s_y^2 \left[1 + b_1 \log\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\right]\left[1 + b_2 \log\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)\right] \tag{2.2}$$

$$T_3 = w_3 s_y^2 \left[1 + \log\left(\frac{s_{x_1}^{2*'}}{s_{x_1}^{2*}}\right)\right]^{c_1} \left[1 + \log\left(\frac{s_{x_2}^{2*'}}{s_{x_2}^{2*}}\right)\right]^{c_2} \tag{2.3}$$

$$T_4 = w_4 s_y^2 \left[1 + d_1 \log\left(\frac{s_{x_1}^{2*'}}{s_{x_1}^{2*}}\right)\right]\left[1 + d_2 \log\left(\frac{s_{x_2}^{2*'}}{s_{x_2}^{2*}}\right)\right] \tag{2.4}$$

where $S_{x_i}^{2*} = a_i S_{x_i}^2 + b_i$ and $s_{x_i}^{2*} = a_i s_{x_i}^2 + b_i$ for i = 1, 2

such that a_i, b_i, c_i and d_i are optimizing scalars or functions of the known parameters of the auxiliary variable x_i's such as the standard deviations S_{x_i}, coefficient of variation C_{x_i}, coefficient of kurtosis b_{2x_i}, coefficient of skewness b_{1x_i} and correlation coefficient $r_{x_i x_j}$ of the population $(i \neq j = 0)$.

3 Properties of the suggested class of estimators

In order to obtain the bias and mean square error (MSE), let us consider

$$\varepsilon_0 = \frac{\left(s_y{}^2 - S_y{}^2\right)}{S_y{}^2},\ \varepsilon_1 = \frac{\left(s_{x_1}{}^2 - S_{x_1}{}^2\right)}{S_{x_1}{}^2},\ \varepsilon_1^{'} = \frac{\left(s_{x_1}{}^{2'} - S_{x_1}{}^2\right)}{S_{x_1}{}^2},\ \varepsilon_2 = \frac{\left(s_{x_2}{}^2 - S_{x_2}{}^2\right)}{S_{x_2}{}^2},\ \varepsilon_2{}' = \frac{\left(s_{x_2}{}^{2'} - S_{x_2}{}^2\right)}{S_{x_2}{}^2}$$

$E(\varepsilon_0)=E(\varepsilon_1)=E(\varepsilon_2)=0$, $E(\varepsilon_0{}^2)=Ib^*_{2y}$, $E(\varepsilon_1{}^2)=Ib^*_{2x_1}$, $E(\varepsilon_1{}^{2'})=I'b^*_{2x_1}$, $E(\varepsilon_2{}^2)=Ib^*_{2x_2}$, $E(\varepsilon_2{}^{2'})=I'b^*_{2x_2}$, $E(\varepsilon_0\varepsilon_1)=I\,I^*_{22yx_1}$, $E(\varepsilon_0\varepsilon_1')=I'I^*_{22yx_1}$, $E(\varepsilon_0\varepsilon_2)=I\,I^*_{22yx_2}$, $E(\varepsilon_0\varepsilon_2')=I'I^*_{22yx_2}$, $E(\varepsilon_1'\varepsilon_2')=I'I^*_{22x_1x_2}$ and $E(\varepsilon_1'\varepsilon_2')=I'I^*_{22x_1x_2}$ where $b^*_{2y}=b_{2y}-1$, $b^*_{2x_1}=b_{2x_1}-1$, $b^*_{2x_2}=b_{2x_2}-1$ and $I^*_{22yx_1}=I_{22yx_1}-1$, $I^*_{22yx_2}=I_{22yx_2}-1$, $I^*_{22x_1x_2}=I_{22x_1x_2}-1$; $I_{pq}=m_{pq}\big/m_{20}^{p/2}\,m_{02}^{q/2}$, $m_{pq}=\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^p\left(X_i-\bar{X}\right)^q\Big/N$, $I=1/N$, $I=1/n'$, $b_{2y}=m_{40}\big/m_{20}^2$, $b_{2x}=m_{04}\big/m_{02}^2$ are the coefficient of kurtosis of y and x respectively.

Theorem 1 *The bias and the mean squared error of the proposed estimator considered upto the terms of order n^{-1} are given by*

$$Bias(T_1)=S_y^2\left[w_1\left\{1+(I-I')\left(\frac{a_1^2}{2}b^*_{2x_1}+\frac{a_2^2}{2}b^*_{2x_2}-a_1r_{yx_1}\sqrt{b^*_{2y}b^*_{2x_1}}-a_2r_{yx_2}\sqrt{b^*_{2y}b^*_{2x_2}}+a_1a_2r_{x_1x_2}\sqrt{b^*_{2x_1}b^*_{2x_2}}\right)\right\}-1\right]$$

$$MSE(T_1)=S_y^4+w_1^4S_y^4\left[1+(I-I')\left\{b^*_{2y}+2a_1^2b^*_{2x_1}+2a_2^2b^*_{2x_2}-4a_1r_{yx_1}\sqrt{b^*_{2y}b^*_{2x_1}}-4a_2r_{yx_2}\sqrt{b^*_{2y}b^*_{2x_2}}+4a_1a_2r_{x_1x_2}\sqrt{b^*_{2x_1}b^*_{2x_2}}\right\}\right]$$
$$-2w_1S_y^4\left\{1+(I-I')\left(\frac{a_1^2}{2}b^*_{2x_1}+\frac{a_2^2}{2}b^*_{2x_2}-a_1r_{yx_1}\sqrt{b^*_{2y}b^*_{2x_1}}-a_2r_{yx_2}\sqrt{b^*_{2y}b^*_{2x_2}}+a_1a_2r_{x_1x_2}\sqrt{b^*_{2x_1}b^*_{2x_2}}\right)\right\}$$

where $r_{yx_1}=\dfrac{I^*_{22yx_1}}{\sqrt{b^*_{2y}b^*_{2x_1}}}$, $r_{yx_2}=\dfrac{I^*_{22yx_2}}{\sqrt{b^*_{2y}b^*_{2x_2}}}$ and $r_{x_1x_2}=\dfrac{I^*_{22x_1x_2}}{\sqrt{b^*_{2x_1}b^*_{2x_2}}}$

Proof. Consider the estimator

$$T_1=w_1s_y^2\left[1+\log\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\right]^{a_1}\left[1+\log\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)\right]^{a_2}$$

$$=w_1S_y^2(1+\varepsilon_0)\left[1+\log(1+\varepsilon_1')(1+\varepsilon_1)^{-1}\right]^{a_1}\left[1+\log(1+\varepsilon_2')(1+\varepsilon_2)^{-1}\right]^{a_2}$$

$$=w_1S_y^2(1+\varepsilon_0)\left[1+a_1\left(\varepsilon_1'-\varepsilon_1-\varepsilon_1\varepsilon_1'+\varepsilon_1^2\right)+\frac{a_1^2}{2}\left(\varepsilon_1'-\varepsilon_1\right)^2-a_1\left(\varepsilon_1'-\varepsilon_1\right)^2\right]$$
$$\left[1+a_2\left(\varepsilon_2'-\varepsilon_2-\varepsilon_2\varepsilon_2'+\varepsilon_2^2\right)+\frac{a_2^2}{2}\left(\varepsilon_2'-\varepsilon_2\right)^2-a_2\left(\varepsilon_2'-\varepsilon_2\right)^2\right] \quad (3.1)$$

On solving and then taking expectation on both the sides, we get

$$Bias(T_1)=S_y^2\left[w_1\left\{1+(I-I')\left(\frac{a_1^2}{2}b^*_{2x_1}+\frac{a_2^2}{2}b^*_{2x_2}-a_1r_{yx_1}\sqrt{b^*_{2y}b^*_{2x_1}}-a_2r_{yx_2}\sqrt{b^*_{2y}b^*_{2x_2}}+a_1a_2r_{x_1x_2}\sqrt{b^*_{2x_1}b^*_{2x_2}}\right)\right\}-1\right] \quad (3.2)$$

Squaring and then taking expectation on both the sides of equation (3.1), we get required expression of MSE.

Corollary 1. *The optimum values of constant are obtained as*

$$w_{1opt} = \frac{B}{A}$$

where

$$A = \left[1 + I b^*_{2y} + (I - I')\left\{+2a_1^2 b^*_{2x_1} + 2a_2^2 b^*_{2x_2} - 4a_1 r_{yx_1}\sqrt{b^*_{2y} b^*_{2x_1}} - 4a_2 r_{yx_2}\sqrt{b^*_{2y} b^*_{2x_2}} + 4a_1 a_2 r_{x_1x_2}\sqrt{b^*_{2x_1} b^*_{2x_2}}\right\}\right]$$

$$B = \left\{1 + (I - I')\left(\frac{a_1^2}{2} b^*_{2x_1} + \frac{a_2^2}{2} b^*_{2x_2} - a_1 r_{yx_1}\sqrt{b^*_{2y} b^*_{2x_1}} - a_2 r_{yx_2}\sqrt{b^*_{2y} b^*_{2x_2}} + a_1 a_2 r_{x_1x_2}\sqrt{b^*_{2x_1} b^*_{2x_2}}\right)\right\}$$

The optimum mean squared error is given by

$$M(T_1)_{opt} = S_y^4\left(1 - \frac{B^2}{A}\right) \tag{3.4}$$

4. Multivariate extension of proposed class of estimators

Let there are k auxiliary variables then we can use the variables by taking a linear combination of these k estimators of the form given in section 2, calculated for every auxiliary variable separately, for estimating the population variance. Then the estimators for population variance will be define as

$$T_1^* = w_1 s_y^2 \ \pi_{i=1}^{k}\left[1 + \log\left(\frac{s_{x_i}^{2'}}{s_{x_i}^2}\right)\right]^{a_i}$$

$$T_2^* = w_2 s_y^2 \ \pi_{i=1}^{k}\left[1 + b_i \log\left(\frac{s_{x_i}^{2'}}{s_{x_i}^2}\right)\right]$$

$$T_3^* = w_3 s_y^2 \ \pi_{i=1}^{k}\left[1 + \log\left(\frac{s_{x_i}^{2*'}}{s_{x_i}^{2*}}\right)\right]^{c_i}$$

$$T_4^* = w_4 s_y^2 \ \pi_{i=1}^{k}\left[1 + d_i \log\left(\frac{s_{x_i}^{2*'}}{s_{x_i}^{2*}}\right)\right]$$

where a_i, b_i, c_i and d_i are the optimizing scalars i = 1, 2, …, k.

Theorem 2 *The bias and the mean squared error of the proposed estimator considered upto the terms of order n^{-1} are given by*

$$Bias\left(T_1\right)=S_y^2\left[w_1\left\{1+\left(I-I'\right)\left(\sum_{i=1}^{k}\frac{a_i^2}{2}b^*_{2x_i}-\sum_{i=1}^{k}a_i\,r_{yx_i}\sqrt{b^*_{2y}\,b^*_{2x_i}}+\sum_{i\neq j=1}^{k}a_i\,a_j r_{x_ix_j}\sqrt{b^*_{2x_i}b^*_{2x_j}}\right)\right\}-1\right]$$

$$MSE\left(T_1\right)=S_y^4+w_1^4\,S_y^4\left[1+\left(I-I'\right)\left\{b^*_{2y}+2\sum_{i=1}^{k}a_i^2b^*_{2x_i}-4\sum_{i=1}^{k}a_ir_{yx_i}\sqrt{b^*_{2y}b^*_{2x_1}}+4\sum_{i=1}^{k}a_ia_jr_{x_ix_j}\sqrt{b^*_{2x_i}b^*_{2x_j}}\right\}\right]$$
$$-2w_1\,S_y^4\left\{1+\left(I-I'\right)\left(\sum_{i=1}^{k}\frac{a_i^2}{2}b^*_{2x_i}-\sum_{i=1}^{k}a_i\,r_{yx_i}\sqrt{b^*_{2y}\,b^*_{2x_i}}+\sum_{i\neq j=1}^{k}a_i\,a_j\,r_{x_ix_j}\sqrt{b^*_{2x_i}b^*_{2x_j}}\right)\right\}$$

5 Efficiency comparison

In this section, we compare the proposed classes of estimators with some important estimators. The comparison will be in terms of their MSE up to the order of n^{-1}. The optimum mean squared error of proposed estimator is given by

$$M\left(T_1\right)_{opt}=S_y^4\left(1-\frac{B^2}{A}\right)$$

5.1 General variance estimator

$$\hat{S}_y^2=s_y^2$$

It's mean squared error is given by

$$MSE(\hat{S}_y^2)=S_y^4\,I\,b^*_{2y}>MSE(T_1)_{opt}$$

5.2 The usual ratio type variance estimator

$$\hat{S}_r^{2'}=s_y^2\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)$$

It's mean squared error is given by

$$MSE(\hat{S}_r^{2'})=S_y^4\left[Ib^*_{2y}+\left(I-I'\right)b^*_{2x_1}+b^*_{2x2}-2I^*_{22yx_1}-2I^*_{22yx_2}+2I^*_{22x_1x_2}\right]>MSE(T_1)_{opt}$$

5.3 The product type variance estimator

$$\hat{S}_p^2=s_y^2\left(\frac{s_{x_1}^2}{s_{x_1}^{2'}}\right)\left(\frac{s_{x_2}^2}{s_{x_2}^{2'}}\right)$$

Its mean squared error is given by

$$MSE(\hat{S}_p^{2'})=S_y^4\left[Ib^*_{2y}+\left(I-I'\right)b^*_{2x_1}+b^*_{2x2}+2I^*_{22yx_1}+2I^*_{22yx_2}+2I^*_{22x_1x_2}\right]>MSE(T_1)_{opt}$$

5.5 Singh, Chauhan, Sawan and Smarandache (2011) type variance estimator

$$\hat{S}_s^{2'} = s_y^2 \exp\left(\frac{s_{x_1}^{2'} - s_{x_1}^2}{s_{x_1}^{2'} + s_{x_1}^2}\right)\left(\frac{s_{x_2}^{2'} - s_{x_2}^2}{s_{x_2}^{2'} + s_{x_2}^2}\right)$$

It’s mean squared error is given by

$$MSE(\hat{S}_s^{2'}) = S_y^4 \left[Ib_{2y}^* + (I - I')\left\{\frac{b_{2x_1}^*}{4} + \frac{b_{2x2}^*}{4} - I_{22yx_1}^* - I_{22yx_2}^* + \frac{I_{22x_1x_2}^*}{4}\right\}\right] > MSE(T_1)_{opt}$$

5.6 Olufadi and Kadilar (2014) variance estimator

$$\hat{S}_K^{2'} = s_y^2 \left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)^{a_1} \left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)^{a_2}$$

It’s mean squared error is given by

$$MSE(\hat{S}_K^{2'}) > MSE(T_1)_{opt}$$

5.7 Das and Tripathi (1978) type variance estimator

$$\hat{S}_D^{2'} = s_y^2 \left(\frac{s_{x_1}^{2'}}{s_{x_1}^{2'} + a_1\left(s_{x_1}^2 - s_{x_1}^{2'}\right)}\right)\left(\frac{s_{x_2}^{2'}}{s_{x_2}^{2'} + a_2\left(s_{x_2}^2 - s_{x_2}^{2'}\right)}\right)$$

It’s mean squared error is given by

$$MSE(\hat{S}_D^{2'}) > MSE(T_1)_{opt}$$

6 Empirical study

The data on which we performed the numerical calculation is taken from some natural populations. The source of the data is given as follows.

Population 1. (Chochran, Pg. no. 155). The data concerns about weekly expenditure on food per family.

y : weekly expenditure on food

x_1 : number of persons

x_2 : the weekly family income

Population 2. (Choudhary F. S., Pg. no. 117).

y : area under wheat (in acres) in 1974

ISSN 2229-5518

x_1 : area under wheat (in acres) in 1971

x_2 : area under wheat (in acres) in 1973

The summary and the percent relative efficiency of the following estimators are as follows:

Table 2: **Parameters of the data**

Parameter	Population 1	Population 2
N	33	34
n	11	10
b_{2y}^*	4.032	2.725
$b_{2x_1}^*$	1.388	12.366
$b_{2x_1}^*$	1.143	1.912
$I_{22yx_1}^*$	0.305	0.224
$I_{22yx_2}^*$	1.155	2.104
$I_{22x_1x_2}^*$	0.492	0.152

Table 3: **PRE of the estimators**

Estimator	Pop. 1	Pop. 2
$\hat{S}_y^2$	100	100
$\hat{S}_r^2$	91.627	29.173
$\hat{S}_p^2$	50.236	17.524
$\hat{S}_s^2$	109.836	75.636
$\hat{S}_D^2$	122.595	230.718
$\hat{S}_K^2$	122.595	230.718
$T_{1_{opt}}$	148.776	233.896

7 Conclusion

This work utilizes the two auxiliary information to the study variable under double sampling. It is clear from the comparative study and numerical study that the proposed estimator perform better than conventional estimators viz. variance estimator, ratio type estimator, product type estimator, Das & Tripathi type (1978) estimator, Singh et al. type (2011) estimator, Olufadi and Kadilar type (2014) estimator etc. Hence, the proposed estimators have much more practical utility than the conventional 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. and Kumari, C. (2018). “A new log type estimators for estimating the population variance”, International Journal of Computational and Applied Mathematics (ISSN 1819-4966), 13 (1), 43-54.

[3] Bhushan, S. and Kumari, C. “A Class of Double Sampling Log-Type Estimators for Population Variance Using Two Auxiliary Variable”, International Journal of Applied Engineering Research, Volume 13, Number (13) 2018, pg 11151-11155.

[4] Bhushan, S. and Kumari, C. “A Class of Double Sampling Log-Type Estimators for Population Variance Using Two Auxiliary Variable”, International Journal of Applied Engineering Research, Volume 13, Number (13) 2018, pg 11151-11155.

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

[6] Bhushan S., Gupta R. and Pandey S. K. (2015). Some log-type classes of estimators using auxiliary information.

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

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

[9] Cochran W. G. (1963). Sampling Techniques, Wiley Eastern Private Limited, New Delhi, 307-310.

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

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

[12] Gupta S. and Shabbir J. (2008). Variance estimation in simple random sampling using auxiliary information, Hacettepe Journal of Mathematics and Statistics, 37, 57-67.

[13] Isaki C. T. (1983). Variance estimation using Auxiliary Information, Jour. Amer. Statist. Asssoc., 78, 117-123.

[14] Kadilar C. and Cingi H. (2006)a . Improvement in variance estimation using auxiliary information, Hacettepe Journal of Mathematics and Statistics, 1(35), 111-115.

[15] Kadilar C. and Cingi H. (2006)b . Ratio estimators for population variance in simple and stratied sampling, Applied Mathematics and Computation, 1(73), 1047-1058.

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

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