Journal of Statistics Applications & Probability Letter
An International Journal

http://dx.doi.org/10.18576/jsapl/070104

Double Sampling Log Type Estimators Using Auxiliary Attribute For Population Variance

Shashi Bhushan[1,*] *and Chandni Kumari*[2]

[1] Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow-226017, India.
[2] Department of Statistics, Babasaheb Bhimrao Ambedkar University (A central University) Lucknow, India.

Received: 2 Jul. 2018, Revised: 2 Aug. 2019, Accepted: 27 Aug. 2019
Published online: 1 Jan. 2020

Abstract: This paper introduces some new classes of log type estimators using the double sampling scheme, assuming that the information on an auxiliary attribute related with the study variable is not known. The mean squared error of the proposed classes of estimators has been obtained up to the ?rst order of approximation. An empirical study is given as an illustration.

Keywords: Estimator, Ratio Method of Estimation, Efficiency, Mean Squared Error.

1 Introduction

Double sampling is widely used when the auxiliary information is not available in advance. We take a large sample of size and then we select a random sample of f?s values for estimating the unknown population variance of auxiliary attribute. Then, we select a random sample of size n from remaining units in the population of size N to estimate the sample variance of study variable and auxiliary attribute. This technique was proposed for the ?rst time by [4]. The literature contains a wide range of ratio, product, di?erence and exponential estimators proposed by various renowned authors like [7], [1], [10], [8], [9] , [11], [12], [13], [14] ,Bhushan and Kumari (2018) among others; but the logarithmic estimator is used here with auxiliary attribute for the estimation of population variance. In this paper, we have tried to make use of the logarithmic relationship between the study variable y and the auxiliary attribute f for estimating the population variance. The proposed estimators would work in case when the study variable is logarithmically related to the auxiliary attribute. 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 $x's$ for estimating the population variance. The proposed estimators would work in case when the study variable is logarithmically related to the auxiliary variable.
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 and f denotes the values of the study and auxiliary variable and attribute for the ith unit $(i = 1, 2, ..., N)$, of the population. Further, let $\bar{y}$ and p be the sample means and $s_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1)$ and $s_x^2 = \sum_{i=1}^{n} (f - P)^2/(n-1)$ be the sample variance of the study and auxiliary variables respectively.

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

We suggest the following new classes of log-type estimators for population variance S_y^2

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

* Corresponding author e-mail: shashi.py@gmail.com

$$T_2 = s_y^2 \left[1 + b_1 \, log \left(\frac{s_f^{2'}}{s_f^2} \right) \right] \quad (2)$$

$$T_3 = s_y^2 \left[1 + log \left(\frac{s_f^{*2'}}{s_f^{*2}} \right) \right]^{c_1} \quad (3)$$

$$T_4 = s_y^2 \left[1 + d_1 \, log \left(\frac{s_f^{*2'}}{s_f^{*2}} \right) \right] \quad (4)$$

where $s_f^{*2'} = a_i s_f^2 + b_i$, $s_f^{*2'} = a_i S_f^2 + b_i$, for $i = 1,2$
such that a_i, b_i, c_i and d_i are the optimizing scalars, $a_i (\neq 0)$, b_i are either real numbers or functions of the known parameters of the auxiliary variable f such as the standard deviations S_f , coefficient of variation C_{xi}, coefficient of kurtosis b_{2f}, coefficient of skewness b_{1xi} and correlation coefficient ρ of the population $(i \neq j = 0)$.

3 Properties of the suggested classes of log-type estimators

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

$$\varepsilon_0 = (s_y^2 - S_y^2)/S_y^2,$$
$$\varepsilon_1 = (s_y^2 - S_y^2)/S_y^2,$$
$$\varepsilon_2 = (s_{x_2}^2 - S_{x_2}^2)/S_{x_2}^2.$$
$$E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0,$$
$$E(\varepsilon_0)^2 = Ib_{2y}^*,$$
$$E(\varepsilon_1)^2 = Ib_{2x_1}^*,$$
$$E(\varepsilon_2)^2 = Ib_{2x_2}^*,$$
$$E(\varepsilon_0 \varepsilon_1) = II_{22yx_1}^*,$$
$$E(\varepsilon_0 \varepsilon_2) = II_{22yx_2}^*,$$
$$E(\varepsilon_1 \varepsilon_2) = II_{22x_1x2}^*,$$

where

$$b_{2x_1}^* = b_{2x_1} - 1,$$
$$b_{2x_2}^* = b_{2x_2} - 1,$$
$$b_{2y}^* = b_{2y} - 1,$$

and

$$I_{22x_1x_2}^* = I_{22x_1x_2} - 1,$$
$$I_{22yx_1}^* = I_{22yx_1} - 1,$$
$$I_{22yx_2}^* = I_{22yx_2} - 1;$$
$$I_{pq} = m_{pq}/m_{20}^{\frac{p}{2}} m_{02}^{\frac{q}{2}},$$
$$m_{pq} = \sum_{i=1}^{N} (Y_i - \bar{Y})^p (Y_i - \bar{Y})^q / N,$$
$$I = \frac{1}{N}, \quad b_{2y} = m_{40}/m_{20}^2, \quad b_{2x} = m_{04}/m_{02}^2$$

are the coefficient of kurtosis of y and x respectively.

Theorem 1. *The bias of the proposed estimators are given as*

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

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

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

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

where,

$$\eta = \frac{aS_f^2}{aS_f^2 + b}$$

Theorem 2. *The mean square error of the proposed estimator considered upto the terms of order n^{-1} are given as follows*

$$MSE(T_1) = S_y^4\left[(I - I')\left\{b_{2y}^* + a_i^2 b_{2f}^* - 2a_i I_{22}^*\right\} + I' b_{2y}^*\right] \; fori = 1,2 \tag{5}$$

$$MSE(T_2) = S_y^4\left[(I - I')\left\{b_{2y}^* + a_i^2 \eta^2 b_{2f}^* - 2a_i \eta I_{22}^*\right\} + I' b_{2y}^*\right] \; fori = 3,4 \tag{6}$$

Corollary 1. *The optimum value of constant are as*

$$a_{1opt} = \frac{I_{22}^*}{b_{2f}^*}$$

The optimum mean squared error is given by

$$M(T_1)_{opt} = S_y^4\left[(I - I')\left\{b_{2y}^* - \frac{I_{22}^{2*}}{b_{2f}^*}\right\} + I' b_{2f}^*\right] \tag{7}$$

4 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?s up to the order of n^{-1}. The optimum mean squared error of proposed estimator is given by

$$M(T_1)_{opt} = S_y^4\left[(I - I')\left\{b_{2y}^* - \frac{I_{22}^{2*}}{b_{2f}^*}\right\} + I' b_{2f}^*\right]$$

4.1 Double sampling ratio estimator using auxiliary attribute

$$\hat{S}_r = s_y^2\left[\frac{s_f^{2'}}{s_f^2}\right]$$

$$MSE(\hat{S}_r^2) = S_y^4\left[(I - I')\left\{b_{2y}^* + b_{2f}^* - 2I_{22}^*\right\} + I' b_{2y}^*\right] > MSE(T_1)_{opt}$$

4.2 Double sampling product estimator using auxiliary attribute

$$\hat{S}_p = s_y^2\left[\frac{s_f^2}{s_f^{2'}}\right]$$

$$MSE(\hat{S}_p^2) = S_y^4\left[(I - I')\left\{b_{2y}^* + b_{2f}^* + 2I_{22}^*\right\} + I' b_{2y}^*\right] > MSE(T_1)_{opt}$$

4.3 Linear regression estimator

$$\hat{S}_L^2 = S_y^2 + d_1\left(s_f^2 - s_f^{2'}\right)$$
$$MSE(\hat{S}_L^2)_{opt} = MSE(T_1)_{opt}$$

4.4 Das and Tripathi type double sampling exponential ratio estimator using auxiliary attribute

$$\hat{S}_D = s_y^2\left[\frac{s_f^{2'}}{s_f^{2'} + \beta\left(s_f^2 - s_f^{2'}\right)}\right]$$
$$MSE(\hat{S}_D^2) = MSE(T_1)_{opt}$$

4.5 Prasad and Singh (1992) estimator

$$\hat{S}_{PS} = s_y^2\,\alpha\left[\frac{s_f^{2'}}{s_f^2}\right]$$
$$MSE(\hat{S}_{PS}^2) = MSE(T_1)_{opt}$$

4.6 Garcia and Cebrian (1996) estimator

$$\hat{S}_G = s_y^2\left[\frac{s_f^{2'}}{s_f^2}\right]^b$$
$$MSE(\hat{S}_G^2) = MSE(T_1)_{opt}$$

4.7 Upadhaya and Singh (2001) estimator

$$\hat{S}_U = S_y^2 + \alpha\left(s_f^{2'} - s_f^2\right)$$
$$MSE(\hat{S}_U^2) = MSE(T_1)_{opt}$$

4.8 Yadav and Kadilar (2013) estimator

$$\hat{S}_Y = s_y^2\,exp\left[1 - \gamma\frac{s_f^{2'}}{s_f^{2'} + (\gamma - 1)s_f^2}\right]$$
$$MSE(\hat{S}_Y^2) = MSE(T_1)_{opt}$$

5 Empirical study

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

Population 1. (Singh D and Chaudhary F. S., Pg. no. 141).
y : number of bearing lime trees
f : area under lime (in acres)
Population 2. (Choudhary F. S. and Singh D., Pg. no. 176).
y : number of cows in milk enumerated
f : number of cows in milk in the previous year
Population 3. (Singh S., Pg. no. 324-325).

y : approximate duration of sleep (in minutes)
f : age in years of the persons
The summary and the percent relative efficiency of the following estimators are as follows:

Table 1: *Parameters of the Data*

Parameters	Population 1	Population 2	Population 3
S_y^2	6564586.45	332721.2079	3582.579
S_f^2	1092.1024	281472.7868	85.2367
b_{2y}^*	12.2574	6.2079	1.6678
b_{2f}^*	4.5788	500.43	1.2389
I_{22}^*	6.7126	4.9528	0.9961
C_f	1.4273	0.8276	0.139
ρ	0.9021	0.8933	0.8552
n'	17	17	25
n	9	8	9

Table 2: *PRE of the estimators*

Estimator	Population 1	Population 2	Population 3
$\hat{S}_y^2$	100	100	100
$\hat{S}_r^2$	111.506	112.114	324.801
$\hat{S}_L^2$	76.207	72.252	243.531
$\hat{S}_D^2$	111.506	112.114	324.801
$\hat{S}_{PS}^2$	111.506	112.114	324.801
$\hat{S}_G^2$	111.506	112.114	324.801
$\hat{S}_U^2$	111.506	112.114	324.801
$\hat{S}_Y^2$	111.506	112.114	324.801
T_{1opt}	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 Perspective

We conclude from the above e?ciency comparison that the proposed estimator are better than the considered estimators. This research work is appropriate when the information about auxiliary attribute is not known. Also, from the empirical study it is clear that proposed estimator is equally e?cient than the other considered estimators. Further, this research work will very useful for survey planners. I recommend this work for publication.

References

[1] Bahl S. and Tuteja R. K. (1991). Ratio and Product type exponential estimator, *Info. Optim. Sci.*, Vol. XII(I), 159-163.
[2] Bhushan S., Gupta R. and Pandey S. K. (2015). Some log-type classes of estimators using auxiliary information.
[3] Hidiroglou M. A. and Sarndal C. E. (1998). Use of auxiliary information for two-phase sampling , *Survey Methodology*, 24(1), 11-20.
[4] Neyman J. (1938). Contribution to the theory of sampling human populations, *J. Amer. Stat. Asso.*, 33, 101-116.
[5] Cochran W. G. (1963). Sampling Techniques , *Wiley Eastern Private Limited*, New Delhi, 307-310.
[6] Chaudhury A. (1978). On estimating the variance of a finite population. *Metrika*, 25, 66-67.
[7] Das A. K. and Tripathi T. P. (1978). Use of auxiliary information in estimating the ?nite population variance. *Sankhya*, C(4), 139-148.
[8] Gupta S. and Shabbir J. (2008). Variance estimation in simple random sampling using auxiliary information, *Hacettepe Journal of Mathematics and Statistics*, 37, 57-67.
[9] Isaki C. T. (1983). Variance estimation using Auxiliary Information , *Jour. Amer. Statist. Asssoct*, 78, 117-123.
[10] Kadilar C. and Cingi H. (2006)a. Improvement in variance estimation using auxiliary information, *Hacettepe Journal of Mathematics and Statistics*, 1(35), 111-115.
[11] Kadilar C. and Cingi H. (2006)b. Ratio estimators for population variance in simple and strati?ed sampling, *Applied Mathematics and Computation*, 1(73), 1047-1058.
[12] Sukhatme P. V., Sukhatme B. V., Sukhatme S. and Ashok C. (1984). Sampling Theory of Surveys with Applications , *Iowa State University Press, Ams.*
[13] Swain A. K. P. C. and Mishra G. (1994). Estimation of population variance under unequal probability sampling , *Sankhya*, B (56), 374-384.
[14] Singh, R., Chauhan, P., Sawan, N. & Smarandache, F. (2011), Improved exponential estimator for population variance using two auxiliary variables, *Ital. J. Pure Appl. Math.s* 28, 101?108.