International Journal of Applied Agricultural Research
ISSN 0973-2683 Volume 14, Number 1 (2019) pp. 31-39
© Research India Publications
http://www.ripublication.com

Some New Log Type Class of Double Sampling Estimators

Shashi Bhushan[1]and Raksoni Gupta[2]

[1] *Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, U.P., India.*

[2]*Department of Statistics, National P.G College, Lucknow, U.P., India.*

Abstract

This paper introduces some new classes of log type estimators using the double sampling scheme, assuming that the information on auxiliary variable related with the study variable is not known. The bias and mean square error of the proposed classes of estimators has been obtained up to the first order of approximation. Further, it has been shown that the minimum MSE under the optimum values of the characterising scalars is same as that of the linear regression estimator. It has also been shown that under the estimated values of characterising scalars attains the minimum MSE of the proposed classes. The proposed classes are compared to some commonly used estimators both theoretically as well as empirically.

Key words: Auxiliary variable, bias, mean square error, double sampling.

INTRODUCTION

In sample surveys, the information on an auxiliary variable related to the study variable is required many times to increase the efficiency of the estimator under consideration. When such auxiliary information is lacking, we can use double sampling or two-phase sampling technique provided that such information may be easily and economically obtained. This technique was proposed for the first time by Neyman (1938). In the absence of the knowledge on the population mean of the auxiliary variable, we use double (two-phase) sampling. The double sampling techniques prove to be an economical and powerful technique for finding the reliable estimate in the first phase sample for the unknown parameters of the auxiliary variable x and hence has eminent role to play in survey sampling, for instance, see

Hidiroglou and Sarndal (1998), Bhushan (2013), Singh (2001), Singh and Upadhayay (1995), Singh, Upadhayay and Chandra (2004), Srivastava and Jhajj (1980), etc.

Consider a finite population $U=(U_1,U_2,\ldots,U_N)$ of size N from which a sample of size n is drawn according to simple random sampling without replacement (SRSWOR). Let y_k and x_k denotes the values of the study and auxiliary variables for the k^{th} unit, $(k=1,2,...,N)$, of the population. Further let $\bar{y}$ and $\bar{x}$ be the sample means of the study and auxiliary variables, respectively.

Consider the following estimators of the study variable

(i). General estimator of population mean in case of SRSWOR

$$\bar{Y}_1=\bar{y}= \text{ sample mean}$$

$$\text{with } MSE\left(\bar{Y}_1\right)=f_n\bar{Y}^2C_y^2 \tag{1.1}$$

(ii). Double sampling ratio estimator using auxiliary variable

$$\bar{Y}_2=\frac{\bar{y}}{\bar{x}}\bar{x}'$$

$$\text{with } MSE\left(\bar{Y}_2\right)=\bar{Y}^2\left[f_nC_y^2+f_{nn'}\left(C_x^2-2\rho_{yx}C_yC_x\right)\right] \tag{1.2}$$

(iii). Bahl and Tuteja (1991) double sampling exponential ratio estimator using auxiliary variable

$$\bar{Y}_3=\bar{y}\exp\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}}\right)$$

$$\text{with } MSE\left(\bar{Y}_3\right)=\bar{Y}^2\left[f_nC_y^2+f_{nn'}\left(\frac{1}{4}C_x^2-\rho_{yx}C_yC_x\right)\right] \tag{1.3}$$

Following the log-type estimators proposed by Bhusan, Gupta and Pandey (2015), we propose the following new classes of double sampling based log-type estimators for the population mean $\bar{Y}$:

$$T_1=\bar{y}\left[1+\log\left(\frac{\bar{x}}{\bar{x}'}\right)\right]^{\alpha_1} \tag{1.4}$$

$$T_2=\bar{y}\left[1+\alpha_2\log\left(\frac{\bar{x}}{\bar{x}'}\right)\right] \tag{1.5}$$

$$T_3=\bar{y}\left[1+\log\left(\frac{\bar{x}^*}{\bar{x}'^*}\right)\right]^{\alpha_3} \tag{1.6}$$

$$T_4 = \bar{y}\left[1+\alpha_4 \log\left(\frac{\bar{x}^*}{\bar{x}'^*}\right)\right] \tag{1.7}$$

where

$$\bar{x}'^* = a\bar{x}' + b$$
$$\bar{x}^* = a\bar{x} + b$$

such that $a(\neq 0), b$ are either real numbers or functions of the known parameters of the auxiliary variables x such as the standard deviations S_x , coefficient of variation C_x, coefficient of kurtosis $\beta_2(x)$ and correlation coefficient ρ of the population.

2. BIAS AND MSE(S) OF THE SUGGESTED CLASS OF LOG-TYPE ESTIMATORS BASED ON DOUBLE SAMPLING

Theorem 2.1

Bias of the suggested classes are obtained as

$$Bais(T_1) = f_{nn'}\bar{Y}\alpha_1\left(\rho_{yx}C_yC_x - \frac{C_x^2}{2} + \frac{(\alpha_1-1)}{2}C_x^2\right) \tag{2.1}$$

$$Bais(T_2) = f_{nn'}\bar{Y}\alpha_2\left(\rho_{yx}C_yC_x - \frac{C_x^2}{2}\right) \tag{2.2}$$

$$Bais(T_3) = f_{nn'}\bar{Y}\alpha_3\left(\upsilon\rho_{yx}C_yC_x - \frac{\upsilon^2}{2}C_x^2 + \frac{(\alpha_3-1)\upsilon^2}{2}C_x^2\right) \tag{2.3}$$

$$Bais(T_4) = f_{nn'}\bar{Y}\alpha_4\left(\upsilon\rho_{yx}C_yC_x - \frac{\upsilon^2}{2}C_x^2\right) \tag{2.4}$$

Theorem 2.2

Mean square error of T_i's are given by

$$MSE(T_i) = \bar{Y}^2\left(f_nC_y^2 + \alpha_i^2 f_{nn'}C_x^2 + 2\alpha_i f_{nn'}\rho_{yx}C_yC_x\right) \qquad i = 1,2 \tag{2.5}$$

$$MSE(T_i) = \bar{Y}^2\left(f_nC_y^2 + \alpha_i^2 f_{nn'}\upsilon^2C_x^2 + 2\alpha_i f_{nn'}\upsilon\rho_{yx}C_yC_x\right) \qquad i = 3,4 \tag{2.6}$$

Corollary 2.3

The minimum value of MSE is obtained for the optimum value of α_i given by

$$\alpha_{i(opt)} = -\rho_{yx}\frac{C_y}{C_x} \qquad i = 1,2 \tag{2.7}$$

$$\alpha_{i(opt)} = \frac{-\rho_{yx}}{\upsilon}\frac{C_y}{C_x} \qquad i = 3,4 \tag{2.8}$$

where $\upsilon = a\overline{x}^{'} \big/ \left(a\overline{x}^{'} + b\right)$; $f_n = \left(1/n - 1/N\right)$; $f_{n'} = \left(1/n' - 1/N\right)$ and $f_{nn'} = \left(f_n - f_{n'}\right)$

and the minimum value of mean square error is obtained as

$$MSE_{\min}\left(T_i\right) = \overline{Y}^2 C_y^2 \left(f_n - f_{nn'}\rho_{yx}^2\right) \tag{2.9}$$

3. SOME MEMBERS OF THE CLASS OF ESTIMATORS T_3 & T_4

It can be easily seen that the classes T_3 & 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 variables x such as the standard deviations S_x , coefficient of variation C_x, coefficient of kurtosis $\beta_2(x)$ and correlation coefficient ρ of the population.

Therefore, a wide variety of estimators can be designed using the above known population parameters. Some of them are given below:

Table 1: Some Generalized members of the proposed class of estimators T_3 and T_4

Double sampling based Log-type estimators			
T_3	T_4	a	b
$T_{3_1} = \overline{y}\left[1 + \log\left(\frac{\overline{x}}{\overline{x}^{'}}\right)\right]^{\alpha_3}$	$T_{4_1} = \overline{y}\left[1 + \alpha_4 \log\left(\frac{\overline{x}}{\overline{x}^{'}}\right)\right]$	1	0
$T_{3_2} = \overline{y}\left[1 + \log\left(\frac{\overline{x} + C_x}{\overline{x}^{'} + C_x}\right)\right]^{\alpha_3}$	$T_{4_2} = \overline{y}\left[1 + \alpha_4 \log\left(\frac{\overline{x} + C_x}{\overline{x}^{'} + C_x}\right)\right]$	1	C_x
$T_{3_3} = \overline{y}\left[1 + \log\left(\frac{\beta_2(x).\overline{x} + C_x}{\beta_2(x).\overline{x}^{'} + C_x}\right)\right]^{\alpha_3}$	$T_{4_3} = \overline{y}\left[1 + \alpha_4 \log\left(\frac{\beta_2(x).\overline{x} + C_x}{\beta_2(x).\overline{x}^{'} + C_x}\right)\right]$	$\beta_2(x)$	C_x
$T_{3_4} = \overline{y}\left[1 + \log\left(\frac{C_x.\overline{x} + \beta_2(x)}{C_x.\overline{x}^{'} + \beta_2(x)}\right)\right]^{\alpha_3}$	$T_{4_4} = \overline{y}\left[1 + \alpha_4 \log\left(\frac{C_x.\overline{x} + \beta_2(x)}{C_x.\overline{x}^{'} + \beta_2(x)}\right)\right]$	C_x	$\beta_2(x)$

Double sampling based Log-type estimators

T_3	T_4	a	b
$T_{3_5}=\bar{y}\left[1+\log\left(\frac{\bar{x}+S_x}{\bar{x}'+S_x}\right)\right]^{\alpha_3}$	$T_{4_5}=\bar{y}\left[1+\alpha_4\log\left(\frac{\bar{x}+S_x}{\bar{x}'+S_x}\right)\right]$	1	S_x
$T_{3_6}=\bar{y}\left[1+\log\left(\frac{\beta_1(x).\bar{x}+S_x}{\beta_1(x).\bar{x}'+S_x}\right)\right]^{\alpha_3}$	$T_{4_6}=\bar{y}\left[1+\alpha_4\log\left(\frac{\beta_1(x).\bar{x}+S_x}{\beta_1(x).\bar{x}'+S_x}\right)\right]$	$\beta_1(x)$	S_x
$T_{3_7}=\bar{y}\left[1+\log\left(\frac{\beta_2(x).\bar{x}+S_x}{\beta_2(x).\bar{x}'+S_x}\right)\right]^{\alpha_3}$	$T_{4_7}=\bar{y}\left[1+\alpha_4\log\left(\frac{\beta_2(x).\bar{x}+S_x}{\beta_2(x).\bar{x}'+S_x}\right)\right]$	$\beta_2(x)$	S_x
$T_{3_8}=\bar{y}\left[1+\log\left(\frac{\bar{x}+\rho}{\bar{x}'+\rho}\right)\right]^{\alpha_3}$	$T_{4_8}=\bar{y}\left[1+\alpha_4\log\left(\frac{\bar{x}+\rho}{\bar{x}'+\rho}\right)\right]$	1	ρ
$T_{3_9}=\bar{y}\left[1+\log\left(\frac{\bar{x}+\beta_2(x)}{\bar{x}'+\beta_2(x)}\right)\right]^{\alpha_3}$	$T_{4_9}=\bar{y}\left[1+\alpha_4\log\left(\frac{\bar{x}+\beta_2(x)}{\bar{x}'+\beta_2(x)}\right)\right]$	1	$\beta_2(x)$
$T_{3_{10}}=\bar{y}\left[1+\log\left(\frac{C_x\bar{x}+\rho}{C_x\bar{x}'+\rho}\right)\right]^{\alpha_3}$	$T_{4_{10}}=\bar{y}\left[1+\alpha_4\log\left(\frac{C_x\bar{x}+\rho}{C_x\bar{x}'+\rho}\right)\right]$	C_x	ρ
$T_{3_{11}}=\bar{y}\left[1+\log\left(\frac{\rho\bar{x}+C_x}{\rho\bar{x}'+C_x}\right)\right]^{\alpha_3}$	$T_{4_{11}}=\bar{y}\left[1+\alpha_4\log\left(\frac{\rho\bar{x}+C_x}{\rho\bar{x}'+C_x}\right)\right]$	ρ	C_x
$T_{3_{12}}=\bar{y}\left[1+\log\left(\frac{\beta_2(x).\bar{x}+\rho}{\beta_2(x).\bar{x}'+\rho}\right)\right]^{\alpha_3}$	$T_{4_{12}}=\bar{y}\left[1+\alpha_4\log\left(\frac{\beta_2(x).\bar{x}+\rho}{\beta_2(x).\bar{x}'+\rho}\right)\right]$	$\beta_2(x)$	ρ
$T_{3_{13}}=\bar{y}\left[1+\log\left(\frac{\rho\bar{x}+\beta_2(x)}{\rho\bar{x}'+\beta_2(x)}\right)\right]^{\alpha_3}$	$T_{4_{13}}=\bar{y}\left[1+\alpha_4\log\left(\frac{\rho\bar{x}+\beta_2(x)}{\rho\bar{x}'+\beta_2(x)}\right)\right]$	ρ	$\beta_2(x)$

It is needless to reiterate that for $\alpha_i=+1$, the above estimators behave as product estimators and for $\alpha_i=-1$, as the ratio estimators.

4. CLASS OF ESTIMATORS BASED ON THE ESTIMATED VALUES OF THE CHARACTERISING SCALARS

The optimum values of the characterising scalars are rarely known in practice; hence they can be estimated by their estimators which are based on the sample values. The optimizing values of the characterising scalars can be written as

$$\alpha_{i(opt)} = -\rho_{yx}\frac{C_y}{C_x} = \frac{\mu_{11}}{\mu_{02}}\frac{\overline{X}}{\overline{Y}} = opt(\alpha_i) \qquad i = 1, 2 \tag{4.1}$$

$$\alpha_{i(opt)} = \frac{-\rho_{yx}}{\upsilon}\frac{C_y}{C_x} = \frac{\mu_{11}}{\upsilon\mu_{02}}\frac{\overline{X}}{\overline{Y}} = opt(\alpha_i) \qquad i = 3, 4 \tag{4.2}$$

where $\mu_{ab} = \frac{1}{N}\sum_{k=1}^{N}\left(Y_k - \overline{Y}\right)^a \left(X_k - \overline{X}\right)^b$

we may also take α_i , as the unbiased estimator of $opt(\alpha_i)$ such that

$$\alpha_i = \frac{\mu_{11}}{\mu_{02}}\frac{\overline{x}}{\overline{y}} \qquad i = 1, 2 \tag{4.3}$$

$$\alpha_i = \frac{\mu_{11}}{\upsilon\mu_{02}}\frac{\overline{x}}{\overline{y}} \qquad i = 3, 4 \tag{4.4}$$

where $\mu_{11} = m_{11}$; $\mu_{02} = m_{02}$ are the estimators of the population parameters μ_{11} and μ_{02} respectively such that $m_{ab} = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_k - \overline{y}\right)^a \left(X_k - \overline{x}\right)^b$

Let us take

$$\mu_{11} = \mu_{11}(1+e_2)$$
$$\mu_{02} = \mu_{02}(1+e_3)$$

with $E(e_2) = E(e_3) = 0$ (4.5)

then under the estimated values of characterising scalars α_i 's, the proposed classes of log-type estimators becomes

$$T_1^* = \overline{y}\left[1+\log\left(\frac{\overline{x}}{\overline{x}'}\right)\right]^{\alpha_1} \tag{4.6}$$

$$T_2^* = \overline{y}\left[1+\alpha_2\log\left(\frac{\overline{x}}{\overline{x}'}\right)\right] \tag{4.7}$$

$$T_3^* = \overline{y}\left[1+\log\left(\frac{\overline{x}^*}{\overline{x}'^*}\right)\right]^{\alpha_3} \tag{4.8}$$

$$T_4^* = \bar{y}\left[1+\alpha_4 \log\left(\frac{\bar{x}^*}{\bar{\bar{x}}'^*}\right)\right] \tag{4.9}$$

Theorem 4.1

Using the results given in Sukhatme and Sukhatme (1991), the mean square error of the classes of estimators comes out to be

$$MSE\left(T_i^*\right) = \bar{Y}^2 C_y^2 \left(f_n - f_{nn'} \rho_{yx}^2 \right)$$
$$= MSE_{\min}\left(T_i\right) \qquad \left(i = 1, 2, 3, 4\right) \tag{4.10}$$
$$= M \text{ (say)}$$

5. COMPARISON WITH THE AVAILABLE ESTIMATORS

Let us now compare the proposed classes of estimators with the estimators proposed in section 1.

(i). General estimator of population mean in case of SRSWOR $\bar{Y}_1 = \bar{y}$ vs T_i^*

From (1.1) and (4.10), we get

$$MSE\left(\bar{Y}_1\right) - M = f_{nn'} \bar{Y}^2 \rho_{yx}^2 C_y^2 \geq 0 \tag{5.1}$$

(ii). Double sampling ratio estimator using auxiliary variable $\bar{Y}_2$ vs T_i^*

From (1.2) and (4.10), we get

$$MSE\left(\bar{Y}_2\right) - M = \bar{Y}^2 f_{nn'} \left(C_x - \rho_{yx} C_y\right)^2 \geq 0 \tag{5.2}$$

(iii). Bahl and Tuteja (1991) double sampling exponential ratio estimator using auxiliary variable

$\bar{Y}_3$ vs T_i^*

From (1.3) and (4.10), we get

$$MSE\left(\bar{Y}_3\right) - M = \bar{Y}^2 f_{nn'} \left(\frac{1}{2} C_x - \rho_{yx} C_y\right)^2 \geq 0 \tag{5.3}$$

From (5.1) to (5.3), it can be easily concluded that the proposed classes of estimators using auxiliary information are better than some of the estimators available in literature.

6. EMPIRICAL STUDY

In order to empirically study the performance of the proposed estimators, following two natural populations have been considered:

Population 1: [Source: William G. Cochran (1977), page-34]

$Y =$ Food cost, $X =$ Family income

$\overline{Y} = 27.40$, $\overline{X} = 72.55$, $C_x = 0.146$, $C_y = 0.369$,

$\rho_{yx} = 0.2521$, $n = 15$, $n' = 25$, $N = 33$

Population 2: [Source: Advance data from vital and health statistics, number 347, October 7, 2004(CDC)]

$Y =$ Height of the people, $X =$ Weight of the people

$\overline{Y} = 140.18$, $\overline{X} = 39.63$, $C_x = 0.482337$, $C_y = 0.191654$,

$\rho_{yx} = 0.973$, $n = 18$, $n' = 25$, $N = 36$

Table 6.1: MSE and PRE of the various estimators with respect to the sample mean

Estimators	Population 1		Population 2	
	MSE	PRE	MSE	PRE
$\overline{Y}_1$	3.717245	100	20.04958	100
$\overline{Y}_2$	3.600181	103.2516	36.17606	55.42225
$\overline{Y}_3$	3.552025	104.6514	10.33417	194.0125
T_i^*	**3.543997**	**104.8885**	**9.419932**	**212.8421**

REFERENCES

[1] Bhushan, S. et al. (2013): Improved sampling strategies in finite population, Scholar Press, Germany.

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

[3] Bahl, S. And Tuteja, R.K. (1991): Ratio and Product type exponential estimator, Information and Optimization Sciences, Vol.XII,I, 159-163

[4] Cochran, W. G. (1977): Sampling Techniques, 3rd ed. New York, John Wiley and Sons

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

[6] Sen,A.R. (1978): Estimation of the population mean when the coefficient of variation is known, Comm.Stat.-Theory Methods, A7, 657-672.

[7] Singh,G.N. (2001): On the use of transformed auxiliary variable in the estimation of population mean in two-phase sampling, Statistics in Transition, 5(3), 405-416.

[8] Singh,G.N. and Upadhyaya,L.N. (1995): A class of modified chain type estimators using two auxiliary variables in two-phase sampling, Metron,LIII, 117-125.

[9] Singh,H.P, Upadhyaya, L.N. and Chandra,P. (2004): A general family of estimators for estimating population mean using two auxiliary variables in two-phase sampling, Statistics in transition, 6(7), 1055-1077.

[10] 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.

[11] Upadhyaya,L.N. and Singh,G.N. (2001): Chain type estimators using transformed auxiliary variable in two-phase sampling, Advances in Modeling and Analysis,38,(1-2), 1-10.

[12] Neyman, J. (1938): Contribution to the theory of sampling human populations, J. Amer. Statist. Assoc., 33: 101-116.