International Journal of Scientific Research in ______________________________ **Research Paper**
Mathematical and Statistical Sciences
Volume-5, Issue-5, pp.199-203, October (2018) **E-ISSN:** 2348-4519

Two-phase Ratio-type Estimator of a Finite Population Mean

P. A. Patel[1] and F. H. Shah[2*]

[1]Department of Statistics, Sardar Patel University, Gujarat, India
[2]Department of Statistics, Sardar Patel University, Gujarat, India

**Corresponding author: shah_fagun@yahoo.co.in , Tel.: +919712132255*

Available online at: www.isroset.org

Received:01/Sept/2018, Accepted:11/Oct/2018, Online: 31/Oct/2018

Abstract**--** This paper deals with the estimation of a finite population mean under two-phase sampling in presence of two auxiliary variables. A ratio-type estimator is proposed and its approximate properties are presented. An empirical study is carried out using three natural populations to compare the performance of the proposed estimator with other estimators discussed in this paper.

Keywords**--** *Auxiliary variables, mean square error (MSE), ratio and regression estimators, simulation, two-phase sampling*

I. INTRODUCTION

Sampling has been the most significant phase of conducting a survey. The problem of estimation of the population mean or total of the survey variable, incorporating auxiliary information on one or two variables at estimation stage, in single-phase and two-phase sampling has been considered by many authors. Neyman [1] introduced the concept of two-phase sampling. In two-phase sampling a large sample is selected at first phase to estimate the auxiliary variable and at second stage the variable of interest is estimated. Two-phase sampling is more efficient and more flexible method for estimation. Numbers of estimators are developed in literature where auxiliary information has increased the efficiency of the estimators. Several ratio and regression type estimators are also introduced to estimate the parametric values of the variable of interest. In the cases where the study variable and the auxiliary variable are highly correlated ratio-type estimator have proven to be the most efficient one. Cochran [2] developed ratio and regression estimators under two-phase sampling to estimate the population mean using one auxiliary variable. For further reading see Khoshnevisan [3], Kadilar and Cingi [4], Gupta and Shabbir [5], and references cited in these papers. Many often, extra auxiliary variable z closely related to x but compared to x remotely relate to y is available. For example, x is current value of a characteristic and z is some previous value of the same characteristic in such case two auxiliary variables are used for estimation. Chand [6] extended the concept of Cochran [2] and developed a chain ratio type estimator under two-phase sampling. Kiregyera [7,8] suggested various estimators using two auxiliary variables under two-phase sampling. Nazir et al. [9] developed an estimator for the population mean under stratified sampling using two auxiliary variables. Motivated from the above work, in such situations, we wish to estimate the population mean under two-phase sampling using simple random sampling at both the phases. The research article is aligned as follows: Section I as introduction, Section II as Related work, Section III contains proposed estimator and methodology, Section IV as empirical comparison and Section V as conclusion.

II. RELATED WORK

Consider a finite population of N identifiable units with the study variable y and auxiliary variables x and z. Let $(\overline{Y},\overline{X},\overline{Z})$, (C_y, C_x, C_z) and $(\rho_{yx}, \rho_{yz}, \rho_{xz})$ denote the population means, coefficients of variation and correlation coefficients respectively. We consider the problem of estimation of the population mean $\overline{Y}$, incorporating partial information on x and full information on z at estimation stage, under two-phase sampling in which a preliminary random sample s' of size n' is drawn from the population using simple random sampling without replacement (srswor) and a subsample s of size n is drawn using the same sampling design.

The ratio and regression estimators (Cochran, 1977) of $\overline{Y}$ under two-phase sampling are given by

$$\hat{\overline{\overline{Y}}}_{Rd} = \frac{\overline{y}}{\overline{x}}\overline{x}' \tag{1}$$

and

$$\hat{\overline{\overline{Y}}}_{Regd} = \overline{y} + b_{yx}(\overline{x}' - \overline{x}) \tag{2}$$

where $\overline{x}'$ = the mean of *x* for preliminary sample, $\overline{y}$ ($\overline{x}$) = the mean of *y* (*x*) variable for sub-sample, and b_{yx} = subsample regression coefficient between *y* and *x*.

Chand [6] suggested a chain ratio estimator

$$\hat{\overline{\overline{Y}}}_{c} = \frac{\overline{y}}{\overline{x}} \cdot \frac{\overline{x}'}{\overline{z}'} \cdot \overline{Z} \tag{3}$$

Kiregyera [7,8] , noting that this estimator is not preferable when the regression of *x* on *z* is linear but does not pass through the origin, suggested the ratio-to-regression, ratio-in-regression and regression-in-regression estimators

$$\hat{\overline{\overline{Y}}}_{k1} = \frac{\overline{y}}{\overline{x}}\left[\overline{x}' + b'_{xz}\left(\overline{Z} - \overline{z}'\right)\right] \tag{4}$$

$$\hat{\overline{\overline{Y}}}_{k2} = \overline{y} + b_{yx}\left(\overline{x}'\overline{Z}/\overline{z}' - \overline{x}\right) \tag{5}$$

$$\hat{\overline{\overline{Y}}}_{k3} = \overline{y} + b_{yx}\left[\overline{x}' + b'_{xz}\left(\overline{Z} - \overline{z}'\right) - \overline{x}\right] \tag{6}$$

The approximate MSEs of the estimators given in (1) to (6) are available in Ahmed et al. [10].

III. PROPOSED ESTIMATOR AND METHODOLOGY

The proposed estimator is

$$\hat{\overline{\overline{Y}}}_{prop} = \frac{\bar{y}}{\bar{x}}\bar{x}'\left[\frac{\bar{Z}}{\bar{Z} - r_{yz}\dfrac{c_y}{C_z}(\bar{Z} - \bar{z}')}\right] \tag{7}$$

where $\bar{z}'$ = preliminary sample mean of *z*, c_y = subsample coefficient of variation of *y* and r_{yz} = subsample correlation coefficient between *y* and *z*.

Using the Taylor linearization technique and Equations (10.12) and (10.13) given in Stuart and Ord [11] we obtain the approximate bias and MSE of the proposed estimator as follows. Expanding $\hat{\overline{\overline{Y}}}_{prop}$ around $\overline{Y}$ using the Taylor series expansion, retaining the terms up to order $O(n^{-1})$, we find the variance of $\hat{\overline{\overline{Y}}}_{prop}$ as

$$V\left(\hat{\overline{\overline{Y}}}_{prop}\right) = V(\bar{y}) + \left(-\frac{\bar{Y}}{\bar{X}}\right)^2 V(\bar{x}) + \left(\frac{\bar{Y}}{\bar{X}}\right)^2 V(\bar{x}') + \left(-\frac{\bar{Y}\rho_{xy}C_y}{\bar{Z}C_z}\right)^2 V(\bar{z}')$$

$$-2\frac{\bar{Y}}{\bar{X}}Cov(\bar{y},\bar{x}) + 2\frac{\bar{Y}}{\bar{X}}Cov(\bar{y},\bar{x}') - 2\frac{\bar{Y}}{\bar{Z}}\rho_{yz}\frac{C_y}{C_z}Cov(\bar{y},\bar{z}')$$

$$-2\left(\frac{\bar{Y}}{\bar{X}}\right)^2 Cov(\bar{x},\bar{x}') + 2\frac{\bar{Y}^2}{\bar{X}\bar{Z}}\rho_{yx}\frac{C_y}{C_z}Cov(\bar{x},\bar{z}') - 2\frac{\bar{Y}^2}{\bar{X}\bar{Z}}\rho_{yx}\frac{C_y}{C_z}Cov(\bar{x}',\bar{z}')$$

Inserting variances and covariances under two-phase sampling (Singh et al. [12]) we obtain

$$V\left(\hat{\hat{Y}}_{prop}\right) = \bar{Y}^2\left[f_1\left(C_y^2 + C_x^2 - 2\rho_{xy}C_yC_x\right) - f_2\left(C_x^2 - \rho_{xy} - \rho_{yz}^2C_y^2\right)\right]$$

$$= \bar{Y}^2\left[f_3\left(C_y^2 + C_x^2 - 2\rho_{xy}C_yC_x\right) - f_2\left(1 - \rho_{yz}^2\right)C_y^2\right]$$

where

$$f_1 = n^{-1} - N^{-1}, f_2 = n'^{-1} - N^{-1},\ f_3 = f_1 - f_2$$

Likewise, we can find the bias as

$$Bias\left(\hat{\hat{Y}}_{prop}\right) = \bar{Y}\left[f_3\left(C_x^2 - 2\rho_{yx}C_yC_x\right) + f_2\left\{\left(\rho_{xy}\frac{C_y}{C_z}\right)^2 C_z^2 - \rho_{yz}\frac{C_y}{C_z}\rho_{yz}C_yC_z\right\}\right]$$

$$= \bar{Y}\left[f_3\left(C_x^2 - 2\rho_{yx}C_yC_x\right)\right]$$

The bias is negligible if n is large and relative bias vanishes if $C_x^2 = C_yC_x$ i.e., the regression lines passes through the origin.

IV. EMPERICAL COMPARISION

The approximate variance (Ahmed et al. [9]) of each of the estimator given in (1) to (6) and in (7) are computed using each of the real population described below and the performance of an estimator $\hat{\hat{Y}}$ is calculated in terms of percentage relative efficiency (PRE), $PRE = \left(MSE(\hat{\hat{Y}}_{rd})/MSE(\hat{\hat{Y}}\right) \times 100$

and presented in Table I.

Further, the preceding estimators are compared empirically using a Monte Carlo simulation. For this a preliminary sample of size n' and a subsample of size n is selected using two-phase srswor sampling and the values of each estimator are calculated. This procedure is repeated $M = 5000$ times.

For each estimator $\hat{\hat{Y}}$ the relative bias in percentage (RB %) and relative efficiency in percentage (RE %) are calculated as

$$RB\left(\hat{\hat{Y}}\right) = 100 * \left(\overline{\hat{\hat{Y}}} - \bar{Y}\right)/\bar{Y},$$

$$RE\left(\hat{\hat{Y}}\right) = MSE_{sim}\,(\hat{\hat{Y}}_{Rd})/MSE_{sim}(\hat{\hat{Y}})\ X\ 100$$

where

$$\overline{\hat{\hat{Y}}} = \sum_{j=1}^{M}\hat{\hat{Y}}_j/M \quad \text{and} \quad MSE_{sim}\left(\hat{\hat{Y}}\right) = \sum_{j=1}^{M}\left(\hat{\hat{Y}}_j - Y\right)^2/(M-1)$$

The RB% and RE% are compiled in Table II.

Data set I : Cochran [2] (The observations are replicated 7 times)

y: number of paralytic polio cases in the placebo group

x: number of placebo children

z: number of non-inculcated

$N = 238,\ n' = 80,\ n = 30,\ \bar{Y} = 2.59,\ \bar{X} = 4.92,\ \bar{Z} = 8.37,$
$\rho_{yx} = 0.7328,\quad \rho_{yz} = 0.7290,\ \ \rho_{xz} = 0.9014,\ \ C_y = 0.0546,$
$C_x = 1.0103,\ C_z = 1.015,\ \beta_1(z) = 2.38,\ \beta_2(z) = 8.93$

Data set II : Jobson [13] **(**The observations are replicated 2 times)

y: Highway Rate
x: Weight
z: Engine size

$N = 194, \quad n' = 80, \quad n = 30, \quad \overline{Y} = 68.37, \quad \overline{X} = 2973.71, \quad \overline{Z} = 27.60,$
$\rho_{yx} = 0.7790, \quad \rho_{yz} = 0.7464, \quad \rho_{xz} = 0.8862$
$C_y = 0.1869,\ C_x = 0.1761,\ C_z = 0.4395,\ \beta_1(z) = 0.9441,\ \beta_2(z) = 2.5386$

Data set III: Murthy [14]

y: Number of cultivators in 1961 census
x: Number of households in 1961 census
z: area in sq. miles in 1951 census

$N = 128, n' = 50, n = 20, \bar{Y} = 853.5, \bar{X} = 727.57, \bar{Z} = 5.592$
$\rho_{yx} = 0.7733, \quad \rho_{yz} = 0.6612, \quad \rho_{xz} = 0.5591$
$C_y = 0.6730,\ C_x = 0.5973,\ C_z = 0.6195,\ \beta_1(z) = 0.8840,\ \beta_2(z) = 3.3371$

Table I. PRE of different estimators with respect to $\hat{\hat{Y}}_{rd}$

Estimator	Data I	Data II	Data III
$\hat{\hat{Y}}_{rd}$	100.00	100.00	100.00
$\hat{\hat{Y}}_{regd}$	101.09	103.55	101.62
$\hat{\hat{Y}}_{C}$	96.17	36.75	89.44
$\hat{\hat{Y}}_{k1}$	132.10	133.49	125.43
$\hat{\hat{Y}}_{k2}$	114.67	69.07	157.10
$\hat{\hat{Y}}_{k3}$	133.65	140.31	125.98
$\boldsymbol{\hat{\hat{Y}}_{prop}}$	**304.71**	**293.09**	**440.53**

Table II. RB (%) and RE (%)

Estimator	Relative Bias (%)			Relative Efficiency (%)		
	Population			Population		
	1	2	3	1	2	3
$\hat{\hat{Y}}_{Rd}$	0.28	-0.03	-0.93	100.00	100.00	100.00
$\hat{\hat{Y}}_{Regd}$	0.27	-0.03	-0.87	99.79	99.48	98.03
$\hat{\hat{Y}}_{C}$	0.46	0.08	-0.87	127.91	50.49	114.48
$\hat{\hat{Y}}_{k1}$	0.76	-0.15	-0.82	65.94	55.08	100.99
$\hat{\hat{Y}}_{k2}$	0.45	0.01	-0.86	123.89	68.26	112.12
$\hat{\hat{Y}}_{k3}$	0.52	-0.02	-0.94	123.82	133.99	111.26
$\hat{\hat{Y}}_{prop}$	**0.37**	**-0.06**	**-0.92**	**128.46**	**134.21**	**115.14**

From Tables I and II it is seen that the proposed estimator has reasonable absolute bias and has performed very well for all the populations considered under study.

V. CONCLUSION

Under the assumptions that extra auxiliary variable z closely related to x but compared to x remotely relate to y is available and the relation between the study variable and the main auxiliary variable is straight line passing through the origin a ratio-type estimator is proposed. In such circumstances, our estimator may be more suitable than the estimators included in the paper. This is supported by the empirical study presented above. The estimators discussed in this paper can be further extended in many ways, e.g., using regression method of estimation, estimation of ratio of two study variables, incorporating non-response (Kumar and Kumar [15]) etc.

ACKNOWLEDGMENT

We are grateful to the referees and members of the board of editors for their useful suggestions.

REFERENCES

[1] J. Neyman, "Contribution to the theory of sampling human populations", Journal Amer. Statist. Assoc., 33, pp 101-116, 1938.

[2] W. G. Cochran, *"Sampling Techniques"*, 3rd ed., John Wiley and Sons, New York, 1977.

[3] M. Khoshnevisan, R. Singh, P. Chauhan, N. Sawan, F. Smarandache, *"A general family of estimators for estimating population mean using known value of some population parameter(s)",Far east Journal of Theoretical Statistics* 22, pp.181-191, 2007.

[4] C. Kadilar and H. Cingi, *"Ratio estimators for the population variance in simple and stratified random sampling"*, Appl. Math Computation 173(2), pp.1047-1059, 2006.

[5] S. Gupta, J. Shabbir, *"On the use of transformed auxiliary variables in estimating population mean",* J Stat Plan Inference 137(5), pp.1606-1611, 2007.

[6] L. Chand, *"Some ratio type estimators based on two or more auxiliary variables"*, Unpublished Ph. D. thesis, Iowa State University, Ames, Iowa (USA), 1975.

[7] B. Kiregyera, *"A chain ratio-type estimator in finite population mean in double sampling using two auxiliary variables"*, Metrika 27, pp 217–223, 1980.

[8] B. Kiregyera, *"Regression-type estimator using two auxiliary variables and model of double sampling from finite populations"*, Metrika 31, pp 215–223, 1984.

[9] A.Nazir, R. Jan, S. Maqbool and T.R. Jan, "*A Modified Dual to Ratio Cum Product Estimator Using Two Auxiliary Variables in Stratified Sampling for Estimating Population Mean*", International Journal of Scientific Research in Mathematical and Statistical Sciences, Vol.5, Issue.4, pp.264-271, 2018.

[10] Z. Ahmad, M. Shahbaj, M. Q. and M. Hanif, "*Two Phase Sampling"*, Cambridge Scholars Publishing, UK, pp 62-97, 2013.

[11] A. Stuart and K. Ord, "*Kendall's Advanced Theory of Statistics"*, Volume 1, Distribution Theory, 4th Edition, Wiley Publications, 1987.

[12] H. P. Singh, R.Tailor, and M. S. Kakaran, *"An estimator of Population mean using power transformation"*, J.I.S.A.S., 58(2),pp. 223-230,2004.

[13] J. D. Jobson, "*Applied Multivariate Data Analysis"*, Vol. II, Springer-Verlag, New York, 1992.

[14] M. N. Murthy, "*Sampling theory and methods"*, First edition, Calcutta Statistical Publishing Society, 1967.

[15] K. Kumar, M. Kumar, "*Two Phase Sampling Exponential Type Estimators for Ratio and Product of two Population Means in the Presence of Non-response*", International Journal of Scientific Research in Mathematical and Statistical Sciences, Vol.4, Issue.6, pp.26-34, 2017.

AUTHORS PROFILE

Mr. P. A. Patel completed his M.Sc., M. Phil., Ph. D., from Sardar Patel University, India. He is Professor of Statistics, Department of Statistics, Sardar Patel University, Vallabh Vidyanagar, Gujarat. He specializes in Theory of Sample Surveys and Survey Sampling Inference. He has guided four Ph. D. students and published more than thirty research articles. He has more than thirty three years of teaching experience. Ms. Fagun Shah is currently pursuing her Ph. D. under Mr. P. A. Patel at Sardar Patel University. Her research area is Survey Sampling and shall submit her thesis soon. She has a teaching experience of more than two years with Ahmedabad University.