Olufadi Yunusa

Department of Statistics and Mathematical Sciences

Kwara State University, P.M.B 1530, Malete, Nigeria

Rajesh Singh

Department of Statistics, Banaras Hindu University

Varanasi (U.P.), India

Florentin Smarandache

University of New Mexico, Gallup, USA

Improvement in Estimating The Population Mean Using Dual To Ratio-Cum-Product Estimator in Simple Random Sampling

Published in:
Rajesh Singh, Florentin Smarandache (Editors)
ON IMPROVEMENT IN ESTIMATING POPULATION PARAMETER(S) USING AUXILIARY INFORMATION
Educational Publishing (Columbus) & Journal of Matter Regularity (Beijing), USA - China, 2013
ISBN: 978-1-59973-230-5
pp. 42 - 53

ABSTRACT

In this paper, we propose a new estimator for estimating the finite population mean using two auxiliary variables. The expressions for the bias and mean square error of the suggested estimator have been obtained to the first degree of approximation and some estimators are shown to be a particular member of this estimator. Furthermore, comparison of the suggested estimator with the usual unbiased estimator and other estimators considered in this paper is carried out. In addition, an empirical study with two natural data from literature is used to expound the performance of the proposed estimator with respect to others.

Keywords: Dual-to-ratio estimator; finite population mean; mean square error; multi-auxiliary variable; percent relative efficiency; ratio-cum-product estimator

1. INTRODUCTION

It is well known that the use of auxiliary information in sample survey design results in efficient estimate of population parameters (e.g. mean) under some realistic conditions. This information may be used at the design stage (leading, for instance, to stratification, systematic or probability proportional to size sampling designs), at the estimation stage or at both stages. The literature on survey sampling describes a great variety of techniques for using auxiliary information by means of ratio, product and regression methods. Ratio and product type estimators take advantage of the correlation between the auxiliary variable, x and the study variable, y. For example, when information is available on the auxiliary variable that is positively (high) correlated with the study variable, the ratio method of estimation is a suitable estimator to estimate the population mean and when the correlation is negative the product method of estimation as envisaged by Robson (1957) and Murthy (1964) is appropriate.

Quite often information on many auxiliary variables is available in the survey which can be utilized to increase the precision of the estimate. In this situation, Olkin (1958) was the first author to deal with the problem of estimating the mean of a survey variable when auxiliary variables are made available. He suggested the use of information on more than one supplementary characteristic, positively correlated with the study variable, considering a linear combination of ratio estimators based on each auxiliary variable separately. The coefficients of the linear combination were determined so as to minimize the variance of the estimator. Analogously to Olkin, Singh (1967) gave a multivariate expression of Murthy's (1964) product estimator, while Raj (1965) suggested a method for using multi-auxiliary variables through a linear combination of single difference estimators. More recently, Abu-Dayyeh et al. (2003), Kadilar and Cingi (2004, 2005), Perri (2004, 2005), Dianna and Perri (2007), Malik and Singh (2012) among others have suggested estimators for $\bar{Y}$ using information on several auxiliary variables.

Motivated by Srivenkataramana (1980), Bandyopadhyay (1980) and Singh et al. (2005) and with the aim of providing a more efficient estimator; we propose, in this paper, a new estimator for $\bar{Y}$ when two auxiliary variables are available.

2. BACKGROUND TO THE SUGGESTED ESTIMATOR

Consider a finite population $P = (P_1, P_2, ..., P_N)$ of N units. Let a sample s of size n be drawn from this population by simple random sampling without replacements (SRSWOR). Let y_i and (x_i, z_i) represents the value of a response variable y and two auxiliary variables (x, z) are available. The units of this finite population are identifiable in the sense that they are uniquely labeled from 1 to N and the label on each unit is known. Further, suppose in a survey problem, we are interested in estimating the population mean $\bar{Y}$ of y, assuming that the population means $(\bar{X}, \bar{Z})$ of (x, z) are known. The traditional ratio and product estimators for $\bar{Y}$ are given as

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}} \text{ and } \bar{y}_P = \bar{y}\frac{\bar{z}}{\bar{Z}}$$

respectively, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ are the sample means of y, x and z respectively.

Singh (1969) improved the ratio and product method of estimation given above and suggested the "ratio-cum-product" estimator for $\bar{Y}$ as $\bar{y}_S = \bar{y}\frac{\bar{X}}{\bar{x}}\frac{\bar{z}}{\bar{Z}}$

In literature, it has been shown by various authors; see for example, Reddy (1974) and Srivenkataramana (1978) that the bias and the mean square error of the ratio estimator $\bar{y}_R$, can be reduced with the application of transformation on the auxiliary variable x. Thus, authors like, Srivenkataramana (1980), Bandyopadhyay (1980) Tracy et al. (1996), Singh et al. (1998), Singh et al. (2005), Singh et al. (2007), Bartkus and Plikusas (2009) and Singh et al. (2011) have improved on the ratio, product and ratio-cum-product method of estimation using the transformation on the auxiliary information. We give below the transformations employed by these authors:

$$x_i^* = (1+g)\bar{X} - gx_i \text{ and } z_i^* = (1+g)\bar{Z} - gz_i, \text{ for } i = 1,2,...,N, \quad (1)$$

where $g = \dfrac{n}{N-n}$.

Then clearly, $\bar{x}^* = (1+g)\bar{X} - g\bar{x}$ and $\bar{z}^* = (1+g)\bar{Z} - g\bar{z}$ are also unbiased estimate of $\bar{X}$ and $\bar{Z}$ respectively and $Corr(\bar{y}, \bar{x}^*) = -\rho_{yx}$ and $Corr(\bar{y}, \bar{z}^*) = -\rho_{yz}$. It is to be noted that by using the transformation above, the construction of the estimators for $\bar{Y}$ requires the knowledge of unknown parameters, which restrict the applicability of these estimators. To overcome this restriction, in practice, information on these parameters can be obtained approximately from either past experience or pilot sample survey, inexpensively.

The following estimators $\bar{y}_R^*$, $\bar{y}_P^*$ and $\bar{y}_{SE}$ are referred to as dual to ratio, product and ratio-cum-product estimators and are due to Srivenkataramana (1980), Bandyopadhyay (1980) and Singh et al. (2005) respectively. They are as presented: $\bar{y}_R^* = \bar{y}\dfrac{\bar{x}^*}{\bar{X}}$, $\bar{y}_P^* = \bar{y}\dfrac{\bar{Z}}{\bar{z}^*}$ and $\bar{y}_{SE} = \bar{y}\dfrac{\bar{x}^*}{\bar{X}}\dfrac{\bar{Z}}{\bar{z}^*}$

It is well known that the variance of the simple mean estimator $\bar{y}$, under SRSWOR design is

$$V(\bar{y}) = \lambda S_y^2$$

and to the first order of approximation, the Mean Square Errors (MSE) of $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_S$, $\bar{y}_R^*$, $\bar{y}_P^*$ and $\bar{y}_{SE}$ are, respectively, given by

$$MSE(\bar{y}_R) = \lambda\left(S_y^2 + R_1^2 S_x^2 - 2R_1 S_{yx}\right)$$

$$MSE(\bar{y}_P) = \lambda\left(S_y^2 + R_2^2 S_z^2 + 2R_2 S_{yz}\right)$$

$$MSE(\bar{y}_S) = \lambda\left[S_y^2 - 2D + C\right]$$

$$MSE(\bar{y}_R^*) = \lambda\left(S_y^2 + g^2 R_1^2 S_x^2 - 2gR_1 S_{yx}\right)$$

$$MSE(\bar{y}_P^*) = \lambda\left(S_y^2 + g^2 R_2^2 S_z^2 + 2gR_2 S_{yz}\right)$$

$$MSE(\bar{y}_{SE}) = \lambda\left(S_y^2 + g^2C - 2gD\right)$$

where,

$$\lambda = \frac{1-f}{n}, \quad f = \frac{n}{N}, \quad S_y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2, \quad S_{yx} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right), \quad \rho_{yx} = \frac{S_{yx}}{S_y S_x},$$

$R_1 = \frac{\bar{Y}}{\bar{X}}$, $R_2 = \frac{\bar{Y}}{\bar{Z}}$, $C = R_1^2 S_x^2 - 2R_1 R_2 S_{zx} + R_2^2 S_z^2$, $D = R_1 S_{yx} - R_2 S_{yz}$ and S_j^2 for $(j = x, y, z)$ represents the variances of x, y and z respectively; while S_{yx}, S_{yz} and S_{zx} denote the covariance between y and x, y and z and z and x respectively. Note that ρ_{yz}, ρ_{zx}, S_x^2, S_z^2, S_{yz} and S_{zx} are defined analogously and respective to the subscripts used.

More recently, Sharma and Tailor (2010) proposed a new ratio-cum-dual to ratio estimator of finite population mean in simple random sampling, their estimator with its MSE are respectively given as,

$$\bar{y}_{ST} = \bar{y}\left[\alpha\left(\frac{\bar{X}}{\bar{x}}\right) + (1-\alpha)\left(\frac{\bar{x}^*}{\bar{X}}\right)\right]$$

$$MSE(\bar{y}_{ST}) = \lambda S_y^2\left(1 - \rho_{yx}^2\right).$$

3. PROPOSED DUAL TO RATIO-CUM-PRODUCT ESTIMATOR

Using the transformation given in (1), we suggest a new estimator for $\bar{Y}$ as follows:

$$\bar{y}_{PR} = \bar{y}\left[\theta\left(\frac{\bar{x}^*}{\bar{X}}\frac{\bar{Z}}{\bar{z}^*}\right) + (1-\theta)\left(\frac{\bar{X}}{\bar{x}^*}\frac{\bar{z}^*}{\bar{Z}}\right)\right]$$

We note that when information on the auxiliary variable z is not used (or variable z takes the value `unity') and $\theta = 1$, the suggested estimator $\bar{y}_{PR}$ reduces to the `dual to ratio' estimator $\bar{y}_R^*$ proposed by Srivenkataramana (1980). Also, $\bar{y}_{PR}$ reduces to the `dual to product' estimator $\bar{y}_P^*$ proposed by Bandyopadhyay (1980) estimator if the information on the auxiliary variate x is not used and $\theta = 0$. Furthermore, the suggested estimator reduces to the dual to ratio-cum-product estimator suggested by Singh et al. (2005) when $\theta = 1$ and information on the two auxiliary variables x and z are been utilized.

In order to study the properties of the suggested estimator $\bar{y}_{PR}$ (e.g. MSE), we write

$$\bar{y} = \bar{Y}(1 + k_0);\ \bar{x} = \bar{X}(1 + k_1);\ \bar{z} = \bar{Z}(1 + k_2);$$

with $E(k_0) = E(k_1) = E(k_2) = 0$

and

$$E(k_0^2) = \frac{\lambda S_y^2}{\bar{Y}^2};\quad E(k_1^2) = \frac{\lambda S_x^2}{\bar{X}^2};\quad E(k_2^2) = \frac{\lambda S_z^2}{\bar{Z}^2};\quad E(k_0 k_1) = \frac{\lambda S_{yx}}{\bar{Y}\bar{X}};\quad E(k_0 k_2) = \frac{\lambda S_{yz}}{\bar{Y}\bar{Z}};$$

$$E(k_1 k_2) = \frac{\lambda S_{zx}}{\bar{X}\bar{Z}},$$

Now expressing $\bar{y}_{PR}$ in terms of $k's$, we have

$$\bar{y}_{PR} = \bar{Y}(1 + k_0)\left[\theta(1 - gk_1)(1 - gk_2)^{-1} + (1 - \theta)(1 - gk_1)^{-1}(1 - gk_2)\right] \tag{2}$$

We assume that $|gk_1| < 1$ and $|gk_2| < 1$ so that the right hand side of (2) is expandable.

Now expanding the right hand side of (2) to the first degree of approximation, we have

$$\bar{y}_{PR} - \bar{Y} = \bar{Y}\left[k_0 + (1 - 2\alpha)g(k_1 - k_2 + k_0 k_1 - k_0 k_2) + g^2\left(k_1^2 - k_1 k_2 - \alpha\left(k_1^2 - k_2^2\right)\right)\right] \tag{3}$$

Taking expectations on both sides of (3), we get the bias of $\bar{y}_{PR}$ to the first degree of approximation, as

$$B(\bar{y}_{PR}) = \lambda\bar{Y}\left[gDA + g^2\left(R_1^2 S_x^2 - R_1 R_2 S_{zx} - \theta\left(R_1^2 S_x^2 - R_2^2 S_z^2\right)\right)\right]$$

where $A = 1 - 2\theta$

Squaring both sides of (3) and neglecting terms of $k's$ involving power greater than two, we have

$$\left(\bar{y}_{PR} - \bar{Y}\right)^2 = \bar{Y}^2\left[k_0 + Agk_1 - Agk_2\right]^2$$

$$= \bar{Y}^2\left[k_0^2 + 2Agk_0 k_1 - 2Agk_0 k_2 - 2A^2 g^2 k_1 k_2 + A^2 g^2 k_1^2 + A^2 g^2 k_2^2\right] \tag{4}$$

Taking expectations on both sides of (4), we get the MSE of $\bar{y}_{PR}$, to the first order of approximation, as

$$MSE(\bar{y}_{PR}) = \lambda\left[S_y^2 + 2AgD + A^2g^2C\right] \qquad (5)$$

The MSE of the proposed estimator given in (5) can be re-written in terms of coefficient of variation as

$$MSE(\bar{y}_{PR}) = \lambda\bar{Y}^2\left[C_y^2 + 2AgC_yD^* + A^2g^2C^*\right]$$

where $C^* = C_x^2 + C_z^2 - 2\rho_{zx}C_zC_x$ and $D^* = \rho_{yx}C_x - \rho_{yz}C_z$, $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$, $C_z = \frac{S_z}{\bar{Z}}$

The MSE equation given in (5) is minimized for

$$\theta = \frac{D + Cg}{2Cg} = \theta_0 \text{ (say)}$$

We can obtain the minimum MSE of the suggested estimator $\bar{y}_{PR}$, by using the optimal equation of θ in (5) as follows: $\min . MSE(\bar{y}_{PR}) = \lambda\left[S_y^2 + F(2D + CF)\right]$

where $F = g - E$ and $E = \frac{D + Cg}{C}$

3. EFFICIENCY COMPARISON

In this section, the efficiency of the suggested estimator $\bar{y}_{PR}$ over the following estimator, $\bar{y}$, $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_S$, $\bar{y}_R^*$, $\bar{y}_P^*$, $\bar{y}_{SE}$ and $\bar{y}_{ST}$ are investigated. We will have the conditions as follows:

(a) $MSE(\bar{y}_{PR}) - V(\bar{y}) < 0$ if $\theta > \frac{2D + gC}{2gC}$

(b) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_R) < 0$ if

$Ag(2D + AgC) < R_1\left(R_1S_x^2 - 2S_{yx}\right)$ provided $S_{yx} < \frac{R_1S_x^2}{2}$

(c) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_P) < 0$ if

$$Ag(2D + AgC) < R_2(R_2S_z^2 + 2S_{yz}) \text{ provided } S_{yz} < \frac{R_2S_z^2}{2}$$

(d) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_S) < 0$ if $C < \frac{-2D}{(Ag-1)}$ provided $g < \frac{1}{A}$

(e) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_R^*) < 0$ if

$$4\theta(\theta gC - gC - D) < 2gR_1R_2S_{zx} + 2R_2S_{yz} - 4R_1S_{yx} - gR_2^2S_z^2$$

(f) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_P^*) < 0$ if

$$4\theta(\theta gC - gC - D) < 2gR_1R_2S_{zx} + 4R_2S_{yz} - 2R_1S_{yx} - gR_1^2S_x^2$$

(g) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_{SE}) < 0$ if $g < \frac{-2D}{C(A-1)}$ provided $A < 1$

(h) $MSE(\bar{y}_{PR}) - MSE(\bar{y}_{ST}) < 0$ if $Ag(2D + AgC) < -\rho^2S_y^2$

4. NUMERICAL ILLUSTRATION

To analyze the performance of the suggested estimator in comparison to other estimators considered in this paper, two natural population data sets from the literature are being considered. The descriptions of these populations are given below.

(1) Population I [Singh (1969, p. 377]; a detailed description can be seen in Singh (1965)

y : Number of females employed

x : Number of females in service

z : Number of educated females

$N = 61$, $n = 20$, $\bar{Y} = 7.46$, $\bar{X} = 5.31$, $\bar{Z} = 179$, $S_y^2 = 28.0818$, $S_x^2 = 16.1761$, $S_z^2 = 2028.1953$, $\rho_{xy} = 0.7737$, $\rho_{yz} = -0.2070$, $\rho_{zx} = -0.0033$,

(2) Population II [Source: Johnston 1972, p. 171]; A detailed description of these variables is shown in Table 1.

y : Percentage of hives affected by disease

x : Mean January temperature

z : Date of flowering of a particular summer species (number of days from January 1)

$N = 10$, $n = 4$, $\bar{Y} = 52$, $\bar{X} = 42$, $\bar{Z} = 200$, $S_y^2 = 65.9776$, $S_x^2 = 29.9880$, $S_z^2 = 84$,

$\rho_{xy} = 0.8$, $\rho_{yz} = -0.94$, $\rho_{zx} = -0.073$,

For these comparisons, the Percent Relative Efficiencies (PREs) of the different estimators are computed with respect to the usual unbiased estimator $\bar{y}$, using the formula

$$PRE(., \bar{y}) = \frac{V(\bar{y})}{MSE(.)} \times 100$$

and they are as presented in Table 2.

Table 1: Description of Population II.

y	x	z
49	35	200
40	35	212
41	38	211
46	40	212
52	40	203
59	42	194
53	44	194
61	46	188
55	50	196
64	50	190

Table 2 shows clearly that the proposed dual to ratio-cum-product estimator $\bar{y}_{PR}$ has the highest PRE than other estimators; therefore, we can conclude based on the study populations that the suggested estimator is more efficient than the usual unbiased estimators, the traditional ratio and product estimator, ratio-cum-product estimator by Singh (1969), Srivenkataramana (1980) estimator, Bandyopadhyay (1980) estimator, Singh et al. (2005) estimator and Sharma and Tailor (2010).

Table 2: PRE of the different estimators with respect to $\bar{y}$

Estimators	Population I	Population II
$\bar{y}$	100	100
$\bar{y}_R$	205	277
$\bar{y}_P$	102	187
$\bar{y}_S$	214	395
$\bar{y}_R^*$	215	239
$\bar{y}_P^*$	105	150
$\bar{y}_{SE}$	236	402
$\bar{y}_{ST}$	250	278
$\bar{y}_{PR}$	279	457

5. CONCLUSION

We have developed a new estimator for estimating the finite population mean, which is found to be more efficient than the usual unbiased estimator, the traditional ratio and product estimators and the estimators proposed by Singh (1969), Srivenkataramana (1980), Bandyopadhyay (1980), Singh et al. (2005) and Sharma and Tailor (2010). This theoretical inference is also satisfied by the result of an application with original data. In future, we hope to extend the estimators suggested here for the development of a new estimator in stratified random sampling.

REFERENCES

1. ABU-DAYYEH, W. A., AHMED, M. S., AHMED, R. A., and MUTTLAK, H. A. (2003): Some estimators of finite population mean using auxiliary information, Applied Mathematics and Computation, 139, 287–298.
2. BANDYOPADHYAY, S. (1980): Improved ratio and product estimators, Sankhya series C 42, 45- 49.
3. DIANA, G. and PERRI, P.F. (2007): Estimation of finite population mean using multi-auxiliary information. Metron, Vol. LXV, Number 1, 99-112
4. JOHNSTON, J. (1972): Econometric methods, (2nd edn), McGraw-Hill, Tokyo.
5. KADILAR, C. and CINGI, H. (2004): Estimator of a population mean using two auxiliary variables in simple random sampling, International Mathematical Journal, 5, 357–367.
6. KADILAR, C. and CINGI, H. (2005): A new estimator using two auxiliary variables, Applied Mathematics and Computation, 162, 901–908.
7. Malik, S. and Singh, R. (2012) : Some improved multivariate ratio type estimators using geometric and harmonic means in stratified random sampling. ISRN Prob. and Stat. Article ID 509186, doi:10.5402/2012/509186.
8. MURTHY, M. N. (1964): Product method of estimation. Sankhya, A, 26, 294-307.
9. OLKIN, I. (1958): Multivariate ratio estimation for finite populations, Biometrika, 45, 154–165.
10. PERRI, P. F. (2004): Alcune considerazioni sull'efficienza degli stimatori rapporto-cum-prodotto, Statistica & pplicazioni, 2, no. 2, 59–75.
11. PERRI, P. F. (2005): Combining two auxiliary variables in ratio-cum-product type estimators. Proceedings of Italian Statistical Society, Intermediate Meeting on Statistics and Environment, Messina, 21-23 September 2005, 193–196.
12. RAJ, D. (1965): On a method of using multi-auxiliary information in sample surveys, Journal of the American Statistical Association, 60, 154–165.
13. REDDY, V. N. (1974): On a transformed ratio method of estimation, Sankhya C, **36**, 59–70.
14. SHARMA, B. and TAILOR, R. (2010): A New Ratio-Cum-Dual to Ratio Estimator of Finite Population Mean in Simple Random Sampling. Global Journal of Science Frontier Research, Vol. 10, Issue 1, 27-31
15. SINGH, M.P. (1965): On the estimation of ratio and product of the population parameters, Sankhya series B 27, 321-328.
16. SINGH, M. P. (1967): Multivariate product method of estimation for finite populations, Journal of the Indian Society of Agricultural Statistics, 31, 375–378.
17. SINGH, M.P. (1969): Comparison of some ratio-cum-product estimators, Sankhya series B 31, 375-378.

18. SINGH, H.P., SINGH, R., ESPEJO, M.R., PINEDA, M.D and NADARAJAH, S. (2005): On the efficiency of a dual to ratio-cum-product estimator in sample surveys. Mathematical Proceedings of the Royal Irish Academy, 105A (2), 51-56.
19. SINGH, R., SINGH, H.P AND ESPEJO, M.R. (1998): The efficiency of an alternative to ratio estimator under a super population model. J.S.P.I., 71, 287-301.
20. SINGH, R., CHAUHAN, P. AND SAWAN, N. (2007): On the bias reduction in linear variety of alternative to ratio-cum-product estimator. Statistics in Transition,8(2),293-300.
21. SINGH, R., KUMAR, M., CHAUHAN, P., SAWAN, N. AND SMARANDACHE, F. (2011): A general family of dual to ratio-cum-product estimator in sample surveys. SIT_newseries, 12(3), 587-594.
22. SRIVENKATARAMANA, T. (1978): Change of origin and scale in ratio and difference methods of estimation in sampling, Canad. J. Stat., **6**, 79–86.
23. SRIVENKATARAMANA, T. (1980): A dual to ratio estimator in sample surveys, Biometrika, **67**(1), 199–204.
24. TRACY, D.S., SINGH, H.P. AND SINGH, R. (1996): An alternative to ratio-cum-product estimator in sample surveys. J.S.P.I.,53,375-387.