



A class of almost unbiased modified ratio estimators for population mean with known population parameters

J.Subramani and G.Kumarapandiyam

Department of Statistics, Ramanujan School of Mathematical Sciences Pondicherry University, R V Nagar, Kalapet Puducherry – 605014.

ARTICLE INFO

Article history:

Received: 11 January 2012;

Received in revised form:

13 March 2012;

Accepted: 21 March 2012;

Keywords

Auxiliary variable,
Biased estimators,
Mean squared errors.

ABSTRACT

The present paper deals with estimation of the population mean of the study variable when the information on the auxiliary variable is known and their population parameters are known. In the past, a number of modified ratio estimators are suggested with known values for the Co-efficient of Variation, Co-efficient of Kurtosis, Co-efficient of Skewness, Population Correlation Coefficient etc. However all these modified ratio estimators are biased but with less mean squared errors compared to the usual ratio estimator. In this paper some strategies have been suggested to improve the performance of the existing modified ratio estimators, which lead to a class of almost unbiased modified ratio estimators; and their performances are better than the modified ratio estimators. These are explained with the help of numerical examples.

© 2012 Elixir All rights reserved.

Introduction

In sample surveys, auxiliary information on the finite population under study is quite often available from previous experience, census or administrative databases. The sampling theory describes a wide variety of techniques/ methods for using auxiliary information to improve the sampling design and to obtain more efficient estimators like Ratio, Product and Regression estimators. Ratio estimators, improves the precision of estimate of the population mean or total of a study variable by using prior information on auxiliary variable X which is correlated with the study variable Y . Over the years the ratio method of estimation has been extensively used because of its intuitive appeal and the computational simplicity. Before discussing further about the modified estimators and the proposed estimators the notations to be used are described below:

- N – Population size
- n – Sample size
- Y – Study variable
- X – Auxiliary variable
- $\bar{X}, \bar{Y}$ – Population means
- $\bar{x}, \bar{y}$ – Sample means
- S_x, S_y – Population standard deviations
- C_x, C_y – Coefficient of variations
- ρ – Coefficient of correlation
- $\beta_1 = \frac{N \sum_{i=1}^N (x_i - \bar{x})^3}{(N-1)(N-2)S^3}$ – Coefficient of skewness of the auxiliary variable
- $\beta_2 = \frac{N(N+1) \sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)(N-2)(N-3)S^4} - \frac{3(N-1)^2}{(N-2)(N-3)}$ – Coefficient of kurtosis of the auxiliary variable

- $B(.)$ – Bias of the estimator
- $MSE(.)$ – Mean squared error of the estimator
- $\hat{Y}(\hat{Y}_p)$ – Existing (proposed) modified ratio estimator of $\bar{Y}$

The classical Ratio estimator for the population mean $\bar{Y}$ of the study variable Y is defined as

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}} \bar{X} = \hat{R} \bar{X}, \quad \text{where} \quad \hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{y}{x} \quad (1.1)$$

where $\hat{R}$ is the estimate of $R = \frac{Y}{X} = \frac{y}{x}$, $\bar{y}$ is the sample mean of the study variable Y and $\bar{x}$ is the sample mean of auxiliary variable X . It is assumed that the population mean $\bar{X}$ of auxiliary variable X is known. Among the modified ratio type estimators available in the literature, five of the mostly used modified ratio type estimators are considered for further improvements. For the benefit of the readers the modified ratio estimators together with their biases and their mean squared errors are given below.

When the population coefficient of variation of auxiliary variable C_x is known, Sisodia and Dwivedi [2] has suggested a modified ratio estimator for $\bar{Y}$ together with its bias and mean squared error as given below:

$$\hat{Y}_1 = \bar{y} \left[\frac{\bar{X} + C_x}{\bar{x} + C_x} \right]$$

$$B(\hat{Y}_1) = \frac{(1-f)}{n} \bar{Y} (\theta_1^2 C_x^2 - \theta_1 C_x C_y \rho)$$

$$MSE(\hat{Y}_1) = \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_1^2 C_x^2 - 2\theta_1 C_x C_y \rho)$$

where $f = \frac{n}{N}$ and $\theta_1 = \frac{\bar{x}}{\bar{x} + C_x}$ (1.2)

Motivated by Sisodia and Dwivedi [2], Singh and Kakran [3] has developed another ratio type estimator for $\bar{Y}$ using Coefficient of Kurtosis together with its bias and mean squared error as given below:

$$\begin{aligned}\hat{\bar{Y}}_2 &= \bar{y} \left[\frac{\bar{X} + \beta_2}{\bar{x} + \beta_2} \right] \\ B(\hat{\bar{Y}}_2) &= \frac{(1-f)}{n} \bar{Y} (\theta_2^2 C_x^2 - \theta_2 C_x C_y \rho) \\ MSE(\hat{\bar{Y}}_2) &= \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_2^2 C_x^2 - 2\theta_2 C_x C_y \rho) \\ \text{where } f &= \frac{n}{N} \text{ and } \theta_2 = \frac{\bar{x}}{\bar{x} + \beta_2}\end{aligned} \quad (1.3)$$

Yan and Tian [7] suggested a ratio type estimator using the coefficient of skewness of the auxiliary variable together with its bias and mean squared error as given below:

$$\begin{aligned}\hat{\bar{Y}}_3 &= \bar{y} \left[\frac{\bar{X} + \beta_1}{\bar{x} + \beta_1} \right] \\ B(\hat{\bar{Y}}_3) &= \frac{(1-f)}{n} \bar{Y} (\theta_3^2 C_x^2 - \theta_3 C_x C_y \rho) \\ MSE(\hat{\bar{Y}}_3) &= \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_3^2 C_x^2 - 2\theta_3 C_x C_y \rho) \\ \text{where } f &= \frac{n}{N} \text{ and } \theta_3 = \frac{\bar{x}}{\bar{x} + \beta_1}\end{aligned} \quad (1.4)$$

Using the population correlation coefficient between X and Y , Singh and Tailor [5] proposed another ratio type estimator for $\bar{Y}$ together with its bias and mean squared error as given below:

$$\begin{aligned}\hat{\bar{Y}}_4 &= \bar{y} \left[\frac{\bar{X} + \rho}{\bar{x} + \rho} \right] \\ B(\hat{\bar{Y}}_4) &= \frac{(1-f)}{n} \bar{Y} (\theta_4^2 C_x^2 - \theta_4 C_x C_y \rho) \\ MSE(\hat{\bar{Y}}_4) &= \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_4^2 C_x^2 - 2\theta_4 C_x C_y \rho) \\ \text{where } f &= \frac{n}{N} \text{ and } \theta_4 = \frac{\bar{x}}{\bar{x} + \rho}\end{aligned} \quad (1.5)$$

By using the population variance of auxiliary variable X , Singh [4] proposed a modified ratio estimator for $\bar{Y}$ together with its bias and mean squared error as given below:

$$\begin{aligned}\hat{\bar{Y}}_5 &= \bar{y} \left[\frac{\bar{X} + S_x}{\bar{x} + S_x} \right] \\ B(\hat{\bar{Y}}_5) &= \frac{(1-f)}{n} \bar{Y} (\theta_5^2 C_x^2 - \theta_5 C_x C_y \rho) \\ MSE(\hat{\bar{Y}}_5) &= \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_5^2 C_x^2 - 2\theta_5 C_x C_y \rho) \\ \text{where } f &= \frac{n}{N} \text{ and } \theta_5 = \frac{\bar{x}}{\bar{x} + S_x}\end{aligned} \quad (1.6)$$

For want of space and for the convenience of the readers, the estimators, biases, and mean squared errors discussed in equations (1.2) to (1.6) are represented in a class of modified ratio estimators as given below:

$$\begin{aligned}\hat{\bar{Y}}_i &= \bar{y} \left[\frac{\bar{X} + \alpha_i}{\bar{x} + \alpha_i} \right] \quad i = 1, 2, 3, 4, 5 \\ B(\hat{\bar{Y}}_i) &= \frac{(1-f)}{n} \bar{Y} (\theta_i^2 C_x^2 - \theta_i C_x C_y \rho) \quad i = 1, 2, 3, 4, 5 \\ MSE(\hat{\bar{Y}}_i) &= \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_i^2 C_x^2 - 2\theta_i C_x C_y \rho) \quad i = 1, 2, 3, 4, 5\end{aligned}$$

where $\alpha_1 = C_x$, $\alpha_2 = \beta_2$, $\alpha_3 = \beta_1$, $\alpha_4 = \rho$ and $\alpha_5 = S_x$ (1.7)

$$f = \frac{n}{N}, \theta_1 = \frac{\bar{x}}{\bar{x} + C_x}, \theta_2 = \frac{\bar{x}}{\bar{x} + \beta_2}, \theta_3 = \frac{\bar{x}}{\bar{x} + \beta_1}, \theta_4 = \frac{\bar{x}}{\bar{x} + \rho} \text{ and } \theta_5 = \frac{\bar{x}}{\bar{x} + S_x}$$

The estimators discussed above are biased but having minimum mean squared error compared to ratio estimator. These points have motivated us to introduce a class of almost unbiased modified ratio estimators based on the above estimators. In fact the proposed estimators are all unbiased if the known population parameters are the true values. However in practical problems the known values are replaced by the values estimated from the previous studies or from another sample. Hence these values are not exactly equal to the true value of the population parameters. That is why the proposed estimators are called as a class of almost unbiased modified ratio estimators. Further, the corresponding mean squared errors are lesser than the mean squared errors of the above estimators defined in (1.7). Further the efficiencies of the proposed estimators with that of the modified ratio estimators are assessed for a certain known populations.

Proposed Modified Ratio Estimators

New estimators are generally proposed or constructed by modifying the structure of the sampling designs or the structure of the estimators itself with reasonable and convincing motivations. Moving along this direction, we intend in this paper to show how the problem of estimating the unknown population mean of a study variable can be treated in a unified way by defining a class of estimators which will be (almost) unbiased and more efficient estimators.

The proposed modified ratio estimator for population mean $\bar{Y}$ is

$$\hat{\bar{Y}}_{pi} = k_i \bar{y} \left[\frac{\bar{x} + \alpha_i}{\bar{x} + \alpha_i} \right] \quad i = 1, 2, 3, 4, 5 \quad (2.1)$$

where $k_i = \frac{S_y}{S_y + \lambda_i C_y}$, $\alpha_1 = C_x$, $\alpha_2 = \beta_2$, $\alpha_3 = \beta_1$, $\alpha_4 = \rho$ and $\alpha_5 = S_x$

and λ_i 's are constants. S_y and C_y are the Population variance and Coefficient of variation of study variable Y respectively. It is reasonable to assume that the values of S_y and C_y are known from the previous studies. Then the expected values of the proposed estimators are obtained as

$$E(\hat{\bar{Y}}_{pi}) = E \left(k_i \bar{y} \left[\frac{\bar{x} + \alpha_i}{\bar{x} + \alpha_i} \right] \right) \quad (2.2)$$

where k_i and α_i are as defined above. If we assume that $\lambda_i = 0$ then the proposed estimators are exactly equal to the estimators given in (1.2) to (1.6). If we assume that $\lambda_i = B(\hat{\bar{Y}}_i)$ then the proposed estimators are almost unbiased ratio estimators corresponding to the estimators given in (1.2) to (1.6). The corresponding mean squared errors of the proposed estimators are as given below:

$$MSE(\hat{\bar{Y}}_{pi}) = k_i^2 MSE(\hat{\bar{Y}}_i) \quad i = 1, 2, 3, 4, 5 \quad (2.3)$$

Efficiency Comparison

To the first degree of approximation, the biases of $\hat{\bar{Y}}_i$ ($i=1, 2, 3, 4, 5$) are respectively given by

$$B(\hat{\bar{Y}}_i) = \frac{(1-f)}{n} \bar{Y} (\theta_i^2 C_x^2 - \theta_i C_x C_y \rho) \quad i = 1, 2, 3, 4, 5 \quad (3.1)$$

Similarly the mean squared error of $\hat{\bar{Y}}_i$ ($i=1, 2, 3, 4, 5$) are respectively given by

$$MSE(\hat{\bar{Y}}_i) = \frac{(1-f)}{n} \bar{Y}^2 (C_y^2 + \theta_i^2 C_x^2 - 2\theta_i C_x C_y \rho) \quad i = 1, 2, 3, 4, 5 \quad (3.2)$$

where

$$f = \frac{n}{N}, \theta_1 = \frac{x}{x+C_x}, \theta_2 = \frac{x}{x+\beta_2}, \theta_3 = \frac{x}{x+\beta_1}, \theta_4 = \frac{x}{x+\rho} \text{ and } \theta_5 = \frac{x}{x+S_x}$$

As stated earlier, if $\lambda_i=0$ then the proposed estimator $\hat{\bar{Y}}_{pi}$ reduces to the estimator $\hat{\bar{Y}}_i$. By choosing different values for the constants λ_i one can get a class of estimators for the estimators $\hat{\bar{Y}}_{pi}$. We have the following two situations:

Case -1: When $\lambda_i = B(\hat{\bar{Y}}_i)$

$$\text{In this case } k_i = \frac{S_y}{S_y + B(\hat{\bar{Y}}_i) C_y}$$

By substituting the value of k_i in the proposed estimator and taking expectation we can show that $E[\hat{\bar{Y}}_{pi}] = \bar{Y}$. That is, $\hat{\bar{Y}}_{pi}$ is an unbiased estimator for $\bar{Y}$.

It is true only when the values of the known population parameters are true and exact otherwise the proposed estimators are called as almost unbiased. That is, $\hat{\bar{Y}}_{pi}$ is an almost unbiased estimator of population mean $\bar{Y}$ if the known population parameters are not the true values. When $\lambda_i = B(\hat{\bar{Y}}_i)$ and $B(\hat{\bar{Y}}_i) > 0$ the value of k_i will be less than one, the mean squared errors of the proposed estimator $\hat{\bar{Y}}_{pi}$ will be less than mean squared errors of $\hat{\bar{Y}}_i$.

$$\text{That is, } MSE(\hat{\bar{Y}}_{pi}) < MSE(\hat{\bar{Y}}_i) \quad (3.3)$$

Case- 2: When $0 < \lambda_i < B(\hat{\bar{Y}}_i)$

The proposed estimator $\hat{\bar{Y}}_{pi}$ is an almost unbiased estimator with lesser bias and lesser mean squared error compared to $\hat{\bar{Y}}_i$.

$$\text{That is, } MSE(\hat{\bar{Y}}_{pi}) < MSE(\hat{\bar{Y}}_i) \quad (3.4)$$

Numerical Illustration

The proposed estimators are computed for three populations to demonstrate what we have discussed above. The population 1 is the data given in Khoshnevisan et.al., [6] and the populations 2 and 3 are taken from Murthy [1] in page 228. The population constants obtained from the above data are given below:

Population-1: Khoshnevisan et.al., [6]

N = 20	n = 8	$\bar{Y} = 19.55$	$\bar{X} = 18.8$
$\rho = -0.9199$	$S_y = 6.9441$	$C_y = 0.3552$	$S_x = 7.4128$
$C_x = 0.3943$	$\beta_2 = 3.0613$	$\beta_1 = 0.5473$	$\theta_1 = 0.9794$
$\theta_2 = 0.8599$	$\theta_3 = 0.9717$	$\theta_4 = 1.0514$	$\theta_5 = 0.7172$

Population-2: Murthy [1]

X = Fixed Capital and Y = Output for 80 factories in a region

N = 80	n = 20	$\bar{Y} = 51.8264$	$\bar{X} = 11.2646$
$\rho = 0.9413$	$S_y = 18.3569$	$C_y = 0.3542$	$S_x = 8.4563$
$C_x = 0.7507$	$\beta_2 = -0.06339$	$\beta_1 = 1.05$	$\theta_1 = 0.9375$
$\theta_2 = 1.0056$	$\theta_3 = 0.91473$	$\theta_4 = 0.9228$	$\theta_5 = 0.5712$

Population-3: Murthy [1]

X = Data on number of workers and Y = Output for 80 factories in a region

N = 80	n = 20	$\bar{Y} = 51.8264$	$\bar{X} = 2.8513$
$\rho = 0.9150$	$S_y = 18.3569$	$C_y = 0.3542$	$S_x = 2.7042$
$C_x = 0.9484$	$\beta_2 = 1.3005$	$\beta_1 = 0.6978$	$\theta_1 = 0.7504$
$\theta_2 = 0.6867$	$\theta_3 = 0.8033$	$\theta_4 = 0.7570$	$\theta_5 = 0.5132$

The existing modified ratio estimators with their biases and mean squared errors are given in the following tables:

Conclusion

The present paper has discussed about the strategies for improving the performance of the existing biased ratio type estimators. As a result we have proposed a class of almost unbiased modified ratio estimators based on the available modified ratio estimators and also obtained their mean squared errors. In fact, the proposed estimators are unbiased and more efficient than the existing modified ratio estimators. However the known values of the population parameters are not true and exact to the population parameter values, since more often the known values of the parameters are taken from the previous studies or from another samples obtained for other studies. We support this theoretical result with numerical examples and shown that the proposed class of almost unbiased modified ratio estimators performs better than the existing modified ratio estimators.

Acknowledgement

The second author wishes to record his gratitude and thanks to the Vice Chancellor, Pondicherry University and other University authorities for having given the financial assistance to carry out this research work through the University Fellowship.

References

- [1] Murthy .M.N. (1967): Sampling theory and methods, Statistical Publishing Society, Calcutta ,India.
- [2] Sisodia B.V.S. and Dwivedi V.K. (1981): A modified ratio estimator using coefficient of variation of auxiliary variable. Jour. Ind. Soc. Agri. Stat., 33(1), 13-18.
- [3] Singh H.P. and Kakran M.S. (1993): A modified ratio estimator using known coefficient of kurtosis of an auxiliary character, revised version submitted to Journal of Indian Society of Agricultural Statistics
- [4] Singh G.N. (2003): On the improvement of product method of estimation in sample surveys, Journal of Indian Society of Agricultural Statistics 56 (3), 267-265

[5] Singh H.P. and Tailor R. (2003): Use of known correlation coefficient in estimating the finite population means, Statistics in Transition 6 (4), 555-560

[6] Khoshnevisan M., Singh R., Chauhan P., Sawan N. and Smarandache F. (2007): A general family of estimators for estimating population mean using known value of some

population parameter(s), Far East Journal of Theoretical Statistics 22, 181–191

[7] Yan Z. and Tian B. (2010): Ratio Method to the Mean Estimation Using Coefficient of Skewness of Auxiliary Variable. ICICA 2010, Part II, CCIS 106, pp. 103–110.

Table 4.1: Biases and mean squared errors of the existing modified ratio estimators

Existing Estimators	Population 1		Population 2		Population 3	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_1$	0.4037	15.1265	0.5066	15.2581	0.5360	17.1881
$\hat{\bar{Y}}_2$	0.3310	13.2644	0.6184	19.3382	0.4142	12.8425
$\hat{\bar{Y}}_3$	0.3988	15.0020	0.4714	14.0112	0.6483	21.3660
$\hat{\bar{Y}}_4$	0.4506	16.3099	0.4839	14.4502	0.5496	17.6849
$\hat{\bar{Y}}_5$	0.2527	11.2065	0.0794	2.3565	0.1538	4.7220

The proposed modified ratio estimators together with the values of k_i and the mean squared errors are given below:

Case- 1: When $\lambda_i = B(\hat{\bar{Y}}_i)$

Table 4.2: Values of k_i and mean squared errors of the proposed modified ratio estimators

Proposed Estimators	Population 1		Population 2		Population 3	
	k_i	MSE	k_i	MSE	k_i	MSE
$\hat{\bar{Y}}_{p1}$	0.9797	14.5206	0.9903	14.9641	0.9897	16.8379
$\hat{\bar{Y}}_{p2}$	0.9833	12.8263	0.9882	18.8848	0.9920	12.6397
$\hat{\bar{Y}}_{p3}$	0.9800	14.4081	0.9909	13.7597	0.9876	20.8413
$\hat{\bar{Y}}_{p4}$	0.9774	15.5832	0.9907	14.1841	0.9895	17.3156
$\hat{\bar{Y}}_{p5}$	0.9872	10.9222	0.9984	2.3493	0.9970	4.6940

Case- 2: When $0 < \lambda_i < B(\hat{\bar{Y}}_i)$ and at $\lambda_i = B(\hat{\bar{Y}}_i)/2$

Table 4.3: Values of k_i and mean squared errors of the proposed modified ratio estimators

Proposed Estimators	Population 1		Population 2		Population 3	
	k_i	MSE	k_i	MSE	k_i	MSE
$\hat{\bar{Y}}_{p1}$	0.9897	14.818	0.99513	15.1100	0.9948	17.0116
$\hat{\bar{Y}}_{p2}$	0.9916	13.042	0.99406	19.1095	0.9960	12.7405
$\hat{\bar{Y}}_{p3}$	0.9899	14.700	0.99547	13.8846	0.9937	21.1012
$\hat{\bar{Y}}_{p4}$	0.9886	15.940	0.99535	14.3162	0.9947	17.4988
$\hat{\bar{Y}}_{p5}$	0.9935	11.063	0.99923	2.3529	0.9985	4.7080

From Tables 4.1, 4.2, and 4.3, we summarize the mean squared errors of the existing ($\lambda_i = 0$) and the proposed estimators for different values of λ_i in the tables given below:

Table 4.4: Comparison of Estimators for different values of λ_i for Population-1

Proposed Estimators	$MSE(\hat{\bar{Y}}_{pi})$		$MSE(\hat{\bar{Y}}_{pi})$
	$\lambda_i = 0$ (Existing estimators)	$\lambda_i = B(\hat{\bar{Y}}_i)/2$	$\lambda_i = B(\hat{\bar{Y}}_i)$
$\hat{\bar{Y}}_{p1}$	15.1265	14.8189	14.5206
$\hat{\bar{Y}}_{p2}$	13.2644	13.0426	12.8263
$\hat{\bar{Y}}_{p3}$	15.0020	14.7005	14.4081
$\hat{\bar{Y}}_{p4}$	16.3099	15.9403	15.5832
$\hat{\bar{Y}}_{p5}$	11.2065	11.0630	10.9222

Table 4.5: Comparison of Estimators for different values of λ_i for Population-2

Proposed Estimators	$MSE(\hat{\bar{Y}}_{pi})$		$MSE(\hat{\bar{Y}}_{pi})$
	$\lambda_i = 0$ (Existing estimators)	$\lambda_i = B(\hat{\bar{Y}}_i)/2$	$\lambda_i = B(\hat{\bar{Y}}_i)$
$\hat{\bar{Y}}_{p1}$	15.2581	15.1100	14.9641
$\hat{\bar{Y}}_{p2}$	19.3382	19.1095	18.8848
$\hat{\bar{Y}}_{p3}$	14.0112	13.8846	13.7597
$\hat{\bar{Y}}_{p4}$	14.4502	14.3162	14.1841
$\hat{\bar{Y}}_{p5}$	2.3565	2.3529	2.3493

Table 4.6: Comparison of Estimators for different values of λ_i for Population-3

Proposed Estimators	$MSE(\hat{\bar{Y}}_{pi})$		$MSE(\hat{\bar{Y}}_{pi})$
	$\lambda_i = 0$ (Existing estimators)	$\lambda_i = B(\hat{\bar{Y}}_i)/2$	$\lambda_i = B(\hat{\bar{Y}}_i)$
$\hat{\bar{Y}}_{p1}$	17.1881	17.0116	16.8379
$\hat{\bar{Y}}_{p2}$	12.8425	12.7405	12.6397
$\hat{\bar{Y}}_{p3}$	21.3660	21.1012	20.8413
$\hat{\bar{Y}}_{p4}$	17.6849	17.4988	17.3156
$\hat{\bar{Y}}_{p5}$	4.7220	4.7080	4.6940