Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.11, 2015 www.iiste.org

Almost Unbiased Ratio Estimator Using the Coefficient of variation of unknown auxiliary variable

Mmaduakor Chika[1*] Amahia Godwin[2] Ngwu Benitho[1]

1. Department of Mathematics, Federal University Oye-Ekiti, Nigeria.
2. Department of Statistics, University of Ibadan, Nigeria.

Abstract

Sisodia el at (1981) proposed a modified ratio estimator of population mean when the information on the auxiliary is known. This estimator is biased but with less mean square error compared to the classical ratio estimator. However, the information on the auxiliary variable may not be available in all cases. In this paper, double sampling strategy has been suggested to obtain the information on the auxiliary variable and then the almost unbiased ratio estimator of population mean using coefficient of variation $\hat{C}_x$ has been proposed. To judge the merits of the proposed estimator over other estimators an empirical study is also carried out.

Keywords: Population mean, Coefficient of variation, Double sampling, Bias and Mean squared error.

Introduction

Auxiliary information plays important role in improving the efficiencies of estimators of population parameters, as in ratio, product, regression and difference estimation procedures. In each instant the advance knowledge of the population mean, $\bar{X}$ of the auxiliary variable x is required. In the past, a number modified ratio estimators were suggested with known values for the Co-efficient of variation, Co-efficient of Kurtosis, Co-efficient of Skewness, Population Correlation Coefficient etc. Before discussing further about the modified estimators and the proposed estimator the notations to be used are described below:

$N - Population$

$n - Sample\ size$

$Y -$ 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

$\rho -$ Coefficient of correlation

$\beta_1 = \dfrac{N\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^3}{(N-1)(N-2)S^2} -$ Coefficient of skewness of the auxiliary variable

$\beta_2 = \dfrac{N(N-1)\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^4}{(N-1)(N-2)(N-3)S^2} -$ Coefficient of kurtosis of the auxiliary variable

$B(.) -$ Bias of the estimator

$MSE(.) -$ Mean square error of the estimator

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

$$\hat{\overline{Y}}_{CR} = \frac{\overline{y}}{\overline{x}}\overline{X} = \hat{R}\overline{X} \qquad \text{where } \hat{R} = \frac{\overline{y}}{\overline{x}} = \frac{y}{x} \tag{1.1}$$

Assuming that population mean $\overline{X}$ of the auxiliary x is known, Singh et al. (1993) defined a ratio type estimator using Coefficient of kurtosis together with its bias and mean square error as given below:

$$\hat{\overline{Y}}_{SK} = \overline{y}\left(\frac{\overline{X} + \beta_2}{\overline{x} + \beta_2}\right)$$

$$B\left(\hat{\overline{Y}}_{Sk}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\overline{Y}\left(\theta_1^2 C_x^2 - \theta_1 \rho C_x C_y\right)$$

$$MSE\left(\hat{\overline{Y}}_{Sk}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\overline{Y}^2\left(C_y^2 + \theta_1^2 C_x^2 - 2\theta_1 \rho C_x C_y\right) \quad \text{Where } \theta_1 = \frac{\overline{X}}{\overline{X} + \beta_2} \tag{1.2}$$

Yan et al. (2010) suggested a ratio type estimator using the coefficient of skewness together with its bias and mean square error as given below:

$$\hat{\overline{Y}}_{YK} = \overline{y}\left(\frac{\overline{X} + \beta_1}{\overline{x} + \beta_1}\right)$$

$$B\left(\hat{\overline{Y}}_{YT}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\overline{Y}\left(\theta_2^2 C_x^2 - \theta_2 \rho C_x C_y\right)$$

$$MSE\left(\hat{\overline{Y}}_{YT}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\overline{Y}^2\left(C_y^2 + \theta_2^2 C_x^2 - 2\theta_2 \rho C_x C_y\right) \quad \text{Where } \theta_2 = \frac{\overline{X}}{\overline{X} + \beta_1} \tag{1.3}$$

Using the population correlation between X and Y, Singh et al (2003) proposed another ratio type estimator for $\overline{Y}$ together with its bias and mean square error as given below:

$$\hat{\bar{Y}}_{ST} = \bar{y}\left(\frac{\bar{X}+\rho}{\bar{x}+\rho}\right)$$

$$B\left(\hat{\bar{Y}}_{ST}\right) = \left(\frac{1}{n}-\frac{1}{N}\right)\bar{Y}\left(\theta_3^2 C_x^2 - \theta_3 \rho C_x C_y\right)$$

$$MSE\left(\hat{\bar{Y}}_{ST}\right) = \left(\frac{1}{n}-\frac{1}{N}\right)\bar{Y}^2\left(C_y^2 + \theta_3^2 C_x^2 - 2\theta_3 \rho C_x C_y\right) \quad \text{Where } \theta_3 = \frac{\bar{X}}{\bar{X}+\beta_1} \tag{1.4}$$

The problem of estimating population mean $\bar{Y}$ of y when the population mean $\bar{X}$ is known has been greatly dealt with in the literature. However, in many practical situations when no information is available on the population mean $\bar{X}$ in advance before starting the survey, we estimate $\bar{Y}$ from a sample obtained through a two phase selection. Adopting simple random sampling without replacement (SRSWOR) a large preliminary sample of size n' is selected by SRSWOR from a population of N units. Information on x is obtained from all the n' units and used to estimate $\bar{X}$. A second subsample of size n units $(n < n')$ is selected from the first phase sample units by SRSWOR. Information on both x and y are obtained from this second phase subsample.

The double sampling ratio estimator of the population mean of y is given by

$$\bar{Y}_{dr} = \frac{\bar{y}}{\bar{x}}\bar{x}' = \hat{R}\bar{x}'$$

$$B\left(\bar{Y}_{dr}\right) = \left(\frac{1}{n}-\frac{1}{N}\right)\bar{Y}\left(C_x^2 - \rho C_x C_y\right)$$

$$MSE\left(\bar{Y}_{dr}\right) = \left(\frac{1}{n}-\frac{1}{N}\right)\bar{Y}^2\left(C_y^2 + C_x^2 - 2\rho C_x C_y\right) \tag{1.5}$$

Where

$\bar{x}' = \frac{1}{n}\sum_{i=1}^{n'} x_i$ is the sample estimate of $\bar{X}$ obtained from the first phase sample.

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are the sample estimates of $\bar{Y}$ and $\bar{X}$, respectively, obtained from the second phase sample.

2. Proposed Almost Unbiased Ratio Estimator

Sisodia et al. (1981) proposed a modified ratio estimator for population mean $\bar{Y}$ using information on coefficient of variation of auxiliary variable C_x as

$$\hat{\bar{Y}}_{S} = \bar{y}\left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right) \tag{2.1}$$

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.11, 2015

The estimator $\hat{\hat{Y}}_S$ requires the knowledge of $\overline{X}$. When information is lacking, we define $\hat{\hat{Y}}_S$ in double sampling as

$$\hat{\hat{Y}}_S^{(d)} = \overline{y}\left(\frac{\overline{x}' + C_x}{\overline{x} + C_X}\right) \tag{2.2}$$

To obtain the bias and mean square error of $\hat{\hat{Y}}_S^{(d)}$ we write

$\overline{y} = \overline{Y}(1+e_0)$, $\overline{x} = \overline{X}(1+e_0)$ and $\overline{x}' = \overline{X}(1+e_1)$ such that

$E(e_0) = E(e_1) = E(e_1') = 0$

$E(e_0^2) = f_1 C_y^2, \quad E(e_1') = f_2 C_x^2,\ E(e_1^2) = f_1 C_x^2$

$E(e_0 e_1) = f_1 \rho C_x C_y,\ E(e_0 e_1') = f_2 \rho C_x C_y$ and $E(e_1' e_1) = f_2 C_x^2$

Expressing (2.2) in terms of e_i's we have

$$\hat{\hat{Y}}_S^{(d)} = \overline{Y}(1+e_0)(1+te_1')(1+te_1)^{-1} \tag{2.3}$$

Now using the standard technique, we get bias and mean squared error of the proposed estimator $\hat{\hat{Y}}_S^{(d)}$ up to the first degree of approximation are obtained as

$$B\left(\hat{\hat{Y}}_s^{(d)}\right) = \overline{Y} f_2\left(t^2 C_x^2 - t\rho C_x C_y\right) \tag{2.4}$$

$$MSE\left(\hat{\hat{Y}}_S^{(d)}\right) = \overline{Y}^2\left[f_2 C_y^2 + f_3\left(C_y^2 + C_x^2 - 2\rho C_x C_y\right)\right] \tag{2.5}$$

3. OPTIMUM ALLOCATION AND COMPARISON OF $\hat{\hat{Y}}_S^{(d)}$ WITH SINGLE SAMPLING

Considering the simple cost function,

$$C = c_1 n' + c_2 n\ ;\ c_1 < c_2 \tag{3.1}$$

c_1: is the cost per unit of obtaining information on the auxiliary variable from the first-phase sample.

c_2: is the cost per unit of measuring the study variable from the second-phase sample. Let assume infinite populations for x and y

The MSE of $\hat{\hat{Y}}_S^{(d)}$ in equation (2.5) can be written in this form

$$\hat{\hat{Y}}_S^{(d)} = \frac{1}{n}\overline{Y}^2\left(C_y^2 + t^2 C_x^2 - 2t\rho C_x C_y\right) - \frac{1}{n'}\overline{Y}^2\left(t^2 C_x^2 - 2t\rho C_x C_y\right) \tag{3.2}$$

$$G(n', n, \lambda) = V\left(\hat{\hat{Y}}_S^{(d)}\right) + \lambda\left(c_1 n' + c_2 n - C\right) \tag{3.4}$$

Differentiate G with respect to n', n respectively and equate to zero, we have

$$\frac{\partial G}{\partial n'} = -\frac{1}{n'^2}\overline{Y}^2\left(t^2 C_x^2 - 2t\rho C_x C_y\right) + \lambda c_1 = 0 \tag{3.5}$$

$$\frac{\partial G}{\partial n} = -\frac{1}{n^2}\bar{Y}^2\left(C_y^2 + t^2C_x^2 - 2t\rho C_x C_y\right) + \lambda c_2 = 0 \tag{3.6}$$

Obtain the ratio of equations (3.5) and (3.6) we have,

$$n = n'\left[\frac{C_y^2 + t^2C_x^2 - 2t\rho C_x C_y}{t^2C_x^2 - 2t\rho C_x C_y}\left(\frac{c_1}{c_2}\right)\right]^{1/2} \tag{3.7}$$

Substitute equation (3.7) into the cost function and Optimum values of n' and n become

$$n' = \frac{C}{c_1 + c_2\left[\frac{C_y^2 + t^2C_x^2 - 2t\rho C_x C_y}{t^2C_x^2 - 2t\rho C_x C_y}\left(\frac{c_1}{c_2}\right)\right]^{\frac{1}{2}}} \tag{3.8}$$

$$n = \frac{C}{c_2 + c_1\left[\frac{t^2C_x^2 - t2\rho C_x C_y}{C_y^2 + t^2C_x^2 - 2t\rho C_x C_y}\left(\frac{c_2}{c_1}\right)\right]^{\frac{1}{2}}} \tag{3.9}$$

Substitute the optimum values of n and n' into the MSE of $\hat{\bar{Y}}_S^{(d)}$ to obtain the optimum MES of $\hat{\bar{Y}}_S^{(d)}$ as

$$MSE_0\left(\hat{\bar{Y}}_S^{(d)}\right) = \frac{\bar{Y}^2}{C}\left\{c_2\left(C_y^2 + t^2C_x^2 - t\rho C_x C_y\right) - c_1\left(t^2C_x^2 - 2t\rho C_x C_y\right)\right\} \tag{3.10}$$

The optimum MSE for classical Ratio Estimator $\bar{Y}_{CR}$

$$MES_0\left(\bar{Y}_{CR}\right) = \frac{c_2}{C}\bar{Y}^2\left(C_y^2 + C_x^2 - 2\rho C_x C_y\right) \tag{3.11}$$

3.2 Efficiency Comparison of $\hat{\bar{Y}}_S^{(d)}$

It is observed from (3.10) and (3.11) that the proposed estimator $\hat{\bar{Y}}_S^{(d)}$ is more efficient than Classical ratio Estimator $\bar{Y}_{CR}$ if

$$\rho\frac{C_y}{C_x} > \frac{\left(c_2 - t^2c_2 - t^2c_1\right)}{2\left(c_2 - tc_2 - tc_1\right)} \tag{3.12}$$

4. Empirical study

To analyze the performance of the proposed estimator in comparison to other estimators, one natural population data set is being considered. The population is taken from Murthy (1967) in page 228. The percent relative efficiencies (PREs) of $\hat{\bar{Y}}_R^{(d)}, \hat{\bar{Y}}_{SD}, \hat{\bar{Y}}_{YT}$, $\hat{\bar{Y}}_{SK}$, $\hat{\bar{Y}}_{ST}$ and $\hat{\bar{Y}}_{SD}^{(d)}$ with respect to $\hat{\bar{Y}}_{CR}$. The population constants obtained from the above data are given below:

Population:

X= data on number of workers

Y= Output for 80 factories in a region

N= 80 $n' = 50$ n = 20 $\overline{X} = 3.2618$ $\overline{Y} = 50.8055$

$S_x = 3.2354$ $S_y = 18.3569$ $C_x = 0.9920$ $C_y = 0.3542$

$\beta_2 = 0.4320$ $\beta_1 = -0.0633$ $\rho = 0.9793$

Table -4.1: Percent relative efficiencies of $\hat{\bar{Y}}_R^{(d)}$, $\hat{\bar{Y}}_{SD}$, $\hat{\bar{Y}}_{YT}$, $\hat{\bar{Y}}_{SK}$, $\hat{\bar{Y}}_{ST}$ and $\hat{\bar{Y}}_{SD}^{(d)}$ with respect to $\hat{\bar{Y}}_{CR}$

Estimator	$\hat{\bar{Y}}_{CR}$	$\hat{\bar{Y}}_R^{(d)}$	$\hat{\bar{Y}}_{SD}$	$\hat{\bar{Y}}_{YT}$	$\hat{\bar{Y}}_{SK}$	$\hat{\bar{Y}}_{ST}$	$\hat{\bar{Y}}_{SD}^{(d)}$
PRE	100	105.72	140.80	134.57	131.07	144.85	173.33

Conclusion

Table 4.1 exhibits that the proposed estimators $\hat{\bar{Y}}_{SD}^{(d)}$ is more efficient than classical ratio estimator $\hat{\bar{Y}}_{CR}$, double sampling ratio estimator $\hat{\bar{Y}}_R^{(d)}$ and modified ratio estimators $\hat{\bar{Y}}_{SD}$, $\hat{\bar{Y}}_{YT}$, $\hat{\bar{Y}}_{SK}$, and $\hat{\bar{Y}}_{ST}$.

Larger gain in efficiency is observed by using proposed estimator over other estimators. Therefore, proposed estimator $\hat{\bar{Y}}_{SD}^{(d)}$ is alternative in situations where auxiliary variable is not known in advance.

References

1. 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
2. Murthy M.N. (1967): Sampling theory and methods, Statistics Publishing Society, Calcutta, India.
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. 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.
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.