Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

Simulation Study on Estimation of Mean Per Unit Estimator with Known Coefficient of Variation

Dr. Ashish Kumar Shukla[1], Dr. Sachin Tomer*[1], Dr. Birjesh Kumar[1]

Ramanujan College, University of Delhi

*Corresponding Author Email id: tomerstat@gmail.com

Article History:

Received:12-09-2025

Revised:05-10-2025

Accepted:30-12-2025

Abstract:

One of the aims in survey sampling is to search for the estimators with highest efficiency. In the present paper, an improved estimator over the mean per unit estimator of population mean with known coefficient of variation is proposed. The conditions under which the proposed estimators perform better than the other existing estimators of population mean have been given. A numerical study is also carried out to see the performances of the proposed and existing estimators of population mean and verify the conditions under which proposed estimators are better than other estimators. And a simulation study has been done to verify the proposed estimators perform better than the existing estimators as they are having lesser mean squared error.

Keywords: Coefficient of variation, Bias, Mean square error, Auxiliary variable, Relative Efficiency, Simple Random Sampling

1. Introduction

Sampling is done when the population is very large and we have to get the result very soon. The population parameters are estimated by the corresponding statistics in a natural sense. As it has been mentioned that the most suitable estimator for the estimation of population parameter is the corresponding statistics so to estimate population mean the most suitable estimator is the sample mean. Although he sample mean is an unbiased estimator of population mean and it has reasonably large variance and our aim is to search for the estimator with minimum variance or may be biased but with minimum mean squared error.

Let the variable of interest be y taking the value Y_i for the i^{th} (i=1,2,……….N) unit of the population of size N.
Further let

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \ , \quad \mu_r = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^r , \ C_y = \frac{\sqrt{\mu_2}}{\bar{Y}} = \frac{\sigma_y}{\bar{Y}}, \ \mu_2 = \sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad \text{and} \quad \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

For $y_1, y_2, \ldots\ldots\ldots\ldots y_n$ being the sample observations on y in a simple random sample of size n without replacement, let

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{and} \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$$

For simplicity, it is assumed that the population size N is large enough as compared to the sample size n so that finite population correction (f.p.c.) term may be ignored.

Let

$$\bar{y} = \bar{Y}(1+e_0)\ , \qquad s_y^2 = \sigma_y^2(1+e_1)$$

$$E(e_0) = E(e_1) = 0\ ,\ E(e_1^2) = \frac{\beta_2 - 1}{n}\ , \qquad E(e_0^2) = \frac{\sigma_y^2}{n\bar{Y}^2} = \frac{C_y^2}{n}\ ,\ \ E(e_0 e_1) = \frac{\gamma_1 C_y}{n} = \frac{\mu_3}{n\bar{Y}\sigma_y^2}$$

2. Proposed estimator

Using known coefficient of variation C_y for the estimation of population mean $\bar{Y}$, the proposed estimator is

$$\bar{y}_k = \bar{y} + k\left(\bar{y}^2 - \frac{s_y^2}{C_y^2}\right) \tag{2.1}$$

where k is the characterizing scalar to be chosen suitably.

2.1. Bias and Mean Square Error of Proposed estimator

From (2.1), we have

$$\bar{y}_k = \bar{Y}(1+e_0) + k\left(\bar{Y}^2(1+e_0)^2 - \frac{\sigma_y^2(1+e_1)}{\frac{\sigma_y^2}{\bar{Y}^2}}\right)$$

$$\bar{y}_k = \bar{Y}(1+e_0) + k\left(\bar{Y}^2(1+e_0)^2 - (1+e_1)\right)$$

$$\bar{y}_k = \bar{Y} + \bar{Y}e_0 + k\bar{Y}^2\left(e_0^2 + 2e_0 - e_1\right)$$

$$\bar{y}_k - \bar{Y} = \bar{Y}e_0 + k\bar{Y}^2\left(e_0^2 + 2e_0 - e_1\right) \tag{2.2}$$

Taking expectation on both sides, we have bias up to terms of order O (1/n) to be

$$Bias(\bar{y}_k) = E(\bar{y}_k - \bar{Y}) = k\bar{Y}^2 E(e_0^2) = k\frac{\sigma_y^2}{n} \tag{2.3}$$

Again, squaring both sides of (2.2) and taking expectation, we have mean square error of $\bar{y}_k$ up to terms of order O(1/n) to be

$$\begin{aligned} MSE(\bar{y}_k) &= E(\bar{y}_k - \bar{Y})^2 = E\left\{\bar{Y}e_0 + k\bar{Y}^2\left(e_0^2 + 2e_0 - e_1\right)\right\}^2 \\ &= \bar{Y}^2 E(e_0^2) + k^2\bar{Y}^4 E\left(4e_0^2 + e_1^2 - 4e_0e_1\right) + 2k\bar{Y}^3 E(2e_0^2 - e_0e_1) \\ &= \bar{Y}^2 E(e_0^2) + k^2\bar{Y}^4\left\{4E(e_0^2) + E(e_1^2) - 4E(e_0e_1)\right\} + 2k\bar{Y}^3\left\{2E(e_0^2) - E(e_0e_1)\right\} \end{aligned} \tag{2.4}$$

The optimum value of k minimizing the mean square error of $\bar{y}_k$ in (2.4) is given by

$$k_o = -\frac{\left(2C_y^2 - \gamma_1 C_y\right)}{\left\{4C_y^2 + (\beta_2 - 1) - 4\gamma_1 C_y\right\}} \tag{2.5}$$

and the minimum mean square error of $\bar{y}_k$ is given by

$$MSE(\bar{y}_{k_o}) = \frac{\sigma_y^2}{n} - \frac{\bar{Y}^2\left(2C_y^2 - \gamma_1 C_y\right)^2}{n\left\{4C_y^2 + (\beta_2 - 1) - 4\gamma_1 C_y\right\}} \tag{2.6}$$

3. Estimator with Estimated Optimum $\hat{k}$

For situation when values of β_2 and γ_1 or their good guessed values are not available ,the alternative is to replace these β_2 and γ_1 involved in the optimum k_o by their estimates $\hat{\beta}_2$ and $\hat{\gamma}_1$ based on sample values and get the estimated optimum value of k denoted by $\hat{k}$ as

$$\hat{k} = -\frac{\left(2C_y^2 - \hat{\gamma}_1 C_y\right)}{\left\{4C_y^2 + \left(\hat{\beta}_2 - 1\right) - 4\hat{\gamma}_1 C_y\right\}} \tag{3.1}$$

where

$$\hat{\beta}_2 = \frac{\hat{\mu}_4}{\hat{\mu}_2^{\,2}} \text{ with } \hat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^4 \;,\; \hat{\mu}_2 = s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$$

and $\hat{\gamma}_1 = \frac{\hat{\mu}_3}{\hat{\mu}_2^{\,3/2}}$ with $\hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^3$ and $\hat{\mu}_2^{\,3/2} = s_y^3$

Thus, replacing k by estimated optimum $\hat{k}$ in the estimator $\bar{y}_k$ in (1.1), we get for wider practical utility of the estimator based on the estimated optimum $\hat{k}$ given by

$$\bar{y}_{k_e} = \bar{y} + \hat{k}\left(\bar{y}^2 - \frac{s_y^2}{C_y^2}\right) \tag{3.2}$$

To find the bias and mean square error of $\bar{y}_{k_e}$, let

$\hat{\mu}_3 = \mu_3(1+e_2)$, $\hat{\mu}_4 = \mu_4(1+e_3)$

along with $\bar{y} = \bar{Y}(1+e_0)$ and $s_y^2 = \sigma_y^2(1+e_1)$ so that

$$\hat{k} = -\frac{\left(2C_y^2 - \frac{\mu_3(1+e_2)}{\sigma_y^3(1+e_1)^{3/2}}C_y\right)}{\left[4C_y^2 + \left(\frac{\mu_4(1+e_3)}{\sigma_y^4(1+e_1)^2} - 1\right) - 4\frac{\mu_3(1+e_2)}{\sigma_y^3(1+e_1)^{3/2}}C_y\right]}$$

$$= -\frac{\left[2C_y^2 - \gamma_1\left(1 - \frac{3}{2}e_1 + \frac{15}{8}e_1^2 + e_2 - \frac{3}{2}e_1e_2 - \ldots..\right)C_y\right]}{\left[4C_y^2 + \beta_2\left(1 - 2e_1 + 3e_1^2 + e_3 - 2e_1e_3\right) - 1 - 4\gamma_1\left(1 - \frac{3}{2}e_1 + \frac{15}{8}e_1^2 + e_2 - \frac{3}{2}e_1e_2 - \ldots..\right)C_y\right]}$$

$$= -\frac{\left[2C_y^2 - \gamma_1 C_y\right]}{\left[4C_y^2 + (\beta_2 - 1) - 4\gamma_1 C_y\right]}\left\{1 + \frac{\gamma_1\left(\frac{3}{2}e_1 - \frac{15}{8}e_1^2 - e_2 + \frac{3}{2}e_1e_2 - \ldots\ldots.\right)C_y}{\left(2C_y^2 - \gamma_1 C_y\right)}\right\} *$$

$$\left[1 + \frac{\beta_2\left(-2e_1 + 3e_1^2 + e_3 - 2e_1e_3 - \ldots..\right) + \gamma_1\left(\frac{3}{2}e_1 - \frac{15}{8}e_1^2 - e_2 + \frac{3}{2}e_1e_2 - \ldots...\right)C_y}{\left\{4C_y^2 + (\beta_2 - 1) - 4\gamma_1 C_y\right\}}\right]^{-1}$$

$$= -\frac{\left[2C_y^2-\gamma_1 C_y\right]}{\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]}\left\{1-\frac{\beta_2\left(-2e_1+3e_1^2+e_3-2e_1e_3-.....\right)+\gamma_1\left(\frac{3}{2}e_1-\frac{15}{8}e_1^2-e_2+\frac{3}{2}e_1e_2-......\right)C_y}{\left\{4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right\}}\right\}+$$

$$\frac{\left[2C_y^2-\gamma_1 C_y\right]}{\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]}\left\{\frac{\gamma_1\left(\frac{3}{2}e_1-\frac{15}{8}e_1^2-e_2+\frac{3}{2}e_1e_2-.......\right)C_y}{\left(2C_y^2-\gamma_1 C_y\right)}\right\} \tag{3.3}$$

Substituting $\bar{y}=\bar{Y}(1+e_0)$, $s_y^2=\sigma_y^2(1+e_1)$ and $\hat{k}$ from (3.3) in (3.2), we have

$$\left(\bar{y}_{k_e}-\bar{Y}\right)=\bar{Y}e_0-\frac{\bar{Y}\left(2C_y^2-\gamma_1 C_y\right)}{\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]}\left(2e_0-e_1+e_0^2-e_0e_1+\frac{e_1^2}{2}+........\right) \tag{3.4}$$

Taking expectation of (3.4) and ignoring terms of e's greater than two, we can easily check that the bias of $\bar{y}_{k_e}$ is of order O(1/n); hence , the Bias($\bar{y}_{k_e}$) is negligible for sufficiently large value of n , that is, the estimator $\bar{y}_{k_e}$ is approximately unbiased estimator of the population mean $\bar{Y}$. Further, squaring and taking expectation up to terms of order O(1/n)

$$MSE\left(\bar{y}_{k_e}\right)=E\left[\bar{Y}e_0-\frac{\bar{Y}\left(2C_y^2-\gamma_1 C_y\right)}{n\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]}(2e_0-e_1)\right]^2$$

$$=\frac{\sigma_y^2}{n}-\frac{\bar{Y}^2\left(2C_y^2-\gamma_1 C_y\right)^2}{n\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]} \tag{3.5}$$

which is same as mean square error for the optimum k_o, that is, estimator $\bar{y}_{k_e}$ based on estimated optimum $\hat{k}$ attains the same mean square error as that of the estimator $\bar{y}_{k_o}$ based on optimum k_o.

4. Concluding Remarks

a) For the optimum value k_o of k, it is clear in (2.5) that the estimator $\bar{y}_{k_o}$ attains the minimum mean square error

$$MSE\left(\bar{y}_{k_o}\right)=\frac{\sigma_y^2}{n}-\frac{\bar{Y}^2\left(2C_y^2-\gamma_1 C_y\right)^2}{n\left\{4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right\}} \tag{4.1}$$

b) The estimators $\bar{y}_{k_o}$ with optimum value k_o and the estimator $\bar{y}_{k_e}$ based on estimated optimum $\hat{k}$ have same mean square error given by

$$MSE\left(\bar{y}_{k_e}\right)=MSE\left(\bar{y}_{k_o}\right)$$

$$=\frac{\sigma_y^2}{n}-\frac{\bar{Y}^2\left(2C_y^2-\gamma_1 C_y\right)^2}{n\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]}$$

$$=MSE\left(\bar{y}\right)-\frac{\bar{Y}^2\left(2C_y^2-\gamma_1 C_y\right)^2}{n\left[4C_y^2+(\beta_2-1)-4\gamma_1 C_y\right]} \tag{4.2}$$

which shows that estimators $\bar{y}_{k_e}$ or $\bar{y}_{k_o}$ based on estimated optimum or optimum value are more efficient than the mean per unit estimator $\bar{y}$ in the sense of having lesser mean square error.

c) For normal parent population (that is, for $\gamma_1 = 0$ and $\beta_2 = 3$) , the optimum value k_o from (2.4), reduces to

$$k_o = \frac{C_y^2}{2C_y^2 + 1}$$

for which $MSE(\bar{y}_{k_o})$ becomes

$$MSE(\bar{y}_{k_o}) = \frac{\sigma_y^2}{n} - \frac{2\bar{Y}^2 C_y^4}{n(2C_y^2 + 1)} \tag{4.3}$$

showing that the proposed estimator $\bar{y}_k$ is more efficient than $\bar{y}$ in normal parent population also.

5. Empirical Study

Considering the data given in Cochran (1977, page 34) dealing with the weekly expenditure of family on food (y) group, computation of required values have been done and we have the following

$n = 33$, $\bar{Y} = 27.49$, $\sigma_y^2 = 99.613033$, $C_y = 0.36306$, $\gamma_1 = 1.4651$, $\beta_2 = 5.7146$

Using the required values, we have

$$MSE(\bar{y}) = 3.018576 \tag{5.1}$$

$$MSE(\bar{y}_{k_e}) = MSE(\bar{y}_{k_o}) = 2.4892803 \tag{5.2}$$

From the above, the percent relative efficiency (PRE) of the proposed estimator over the usual mean per unit estimator is 121%.

6. Simulation Study

In this section, simulation is conducted to evaluate the performance of the proposed estimator with respect to traditional estimator. For this study we have generated a population size N = 1000 from standard normal distribution using rnorm package in software R, draw sample of size n = 200. The whole simulation process starting from the drawing sample from variable Y from normal population and calculating the estimate was repeated 50000 times.

$MSE(\bar{y})=0.001300325$\\
$MSE(\bar{y}_{k_e})=MSE(\bar{y}_{k_0 })=0.0009572594$\\

$$MSE(\bar{y}) = 0.001300325 \tag{6.1}$$

$$MSE(\bar{y}_{k_e}) = MSE(\bar{y}_{k_o}) = 0.0009572594 \tag{6.2}$$

From the above, the percent relative efficiency (PRE) of the proposed estimator over the usual mean per unit estimator is 135.8383%.

7. Result and Conclusion

This paper deals with the estimation of population mean of the study variable with known coefficient of variation. The expressions for the bias and mean squared error of the proposed estimator has been derived up to the first order of approximation. A theoretical comparison of the proposed estimators has been made with the existing estimators of population mean under simple random sampling scheme. An empirical study is also carried out to judge the performance of the proposed and existing estimators of population mean. Along with the empirical study a simulation study has been done to check the validity of the proposed estimator. Through this simulation study, it has been found that the proposed estimator is more efficient than the other existing estimator. As proposed estimator are more efficient estimator for population mean.

Refrences

[1] Basu D: "An easy on the logical foundations of survey sampling, Part I.Foundations of Statistical Inference", Ed. By V.P. Godambe and D. A. Sprott. New York, 203-233, (1971).
[2] Cochran, W. G. "Sampling Techniques (Third edition)", John Wiley and Sons Eastern Limited New York, (1977).
SARNDAL
[3] Cassel, C., Sarndal, C., Wretman, J.H., "Foundations of Inference in Survey Sampling", Wiley, New York, (1977).
[4] Cramer, H., "Mathematical Methods of Statistics", Princeton University Press, (1946).
[5] Gleser, L. J., Healy, J. D., "Estimating the mean of a normal distribution with known coefficient of variation", Journal of the American Statistical Association, 71, 977-981, (1976).
[6] Kadilar, C., Cingi, H., "Improvement in estimating the population mean in simple random sampling", Applied Mathematics Letters, 19, 75–79, (2006b).
[7] Mukhopadhyay, P., "Theory and methods of survey sampling", PHI Learning, 2nd edition, New Delhi, (2005).
[8] Meeden G., "Basu's contributions to the foundations of sample survey", In Current Issues in Statistical Inference: Essays in Honor of D. Basu, M. Ghosh, (1992).
[9] Murthy, M. N., "Sampling Theory \& Methods. Statistical Publishing society", Calcutta, (1977).
[10] Searls, D. T., "The Utilization of Known Coefficient of Variation in the Estimation Procedure", J.Amer.Statist. Assoc.59, p1225-26, (1964).
[11] Sen, A. R., "Estimation of the Population Mean When the Coefficient of variation is Known", Communications in Statistics-Theory and Methods, 7, 657-672, (1978).
[12] Singh, H. P., Tailor, R., "Estimation of finite population mean with known co-efficient of variation of an auxiliary variable", Statistica, anno LXV, 3, 301–313, (2005).
[13] Sukhatme, P. V., B. V. Sukhatme, "Sampling theory of surveys with applications (2nd edition)", Asia Publishing House, Bombay, (1970).
[14] Singh, D., Chaudhary, F. S., "Theory and analysis of Sample Surveys Designs" Wiley Eastern Limited New Delhi- (INDIA), (1989).
[15] Sukhatme, P.V., Sukhatme, B. V., Sukhatme, S., Ashok, C., "Sampling Theory of Surveys with Applications", IOWA State University press (AMES), IOWA (U.S.A.) and Indian Society of Agricultural Statistics New Delhi-110012 (India), (1984).
[16] Upadhyaya, L.N., Singh, H.P., "On the Estimation of the Population Mean with Known Coefficient of Variation", Biometrical Journal, 26, 915-922, (1984).