Journal of Reliability and Statistical Studies (ISSN: 0974-8024)
Vol. 1, Issue 1(2008) 42-47

RATIO TYPE ESTIMATOR OF SQUARE OF COEFFICIENT OF VARIATION USING QUALITATIVE AUXILIARY INFORMATION

Sheela Misra[1], Subhash K. Yadav and Anshula Pandey
Dept of Statistics, Lucknow University, Lucknow, India
1. E mail: misra_sheela@yahoo.com

Abstract

This paper deals with the estimation of square of coefficient of variation which is comparatively a more stable quantity using ratio type estimator. Its bias and mean square error (MSE) are found to the first order of approximation. An optimum subclass of estimators is also obtained and a comparative study with the conventional square of sample coefficient of variation estimator is made. It has further been shown that estimation of parametric values involved in the optimum subclass does not reduce the efficiency of the proposed estimator. An empirical example showing the increased efficiency of proposed estimator over square of sample coefficient of variation estimator is also included as an illustration.

Keywords: Qualitative Auxiliary Information, Coefficient of Variation, Estimator, Bias, Mean Square Error, Order of Approximation, Bounds, Finite Population Correction, Efficiency.

1. Introduction

The use of auxiliary information can increase the precision of an estimator when study variable is highly correlated with auxiliary variable x. There exist situations when information is available in the form of attribute ψ which is highly correlated with variable y under study. For example (i) y may be use of drugs and ψ may be the gender (ii) y may be the production of a crop and ψ may be the particular variety.

Let there be N units in the population. Let (Y_i, ψ_i), i = 1,2,…,N be the corresponding observation values of the ith unit of the population of the study variable Y and the auxiliary variable ψ respectively. Further we assume that $\psi_i = 1$ and $\psi_i = 0$, i = 1, 3, …, N depending upon if it possesses a particular characteristic or does not possess it. Let $A = \sum_{i=1}^{N} \psi_i$ and $a = \sum_{i=1}^{n} \psi_i$ denote the total number of units in the population and sample respectively possessing attribute ψ. Let $P = \frac{A}{N}$ and $p = \frac{a}{n}$ denote the proportion of the units in the population and sample respectively possessing attribute ψ. Let a simple random sample of size n from this population is taken without replacement having sample values (y_i, ψ_i) ; i = 1,2, 3, …, n.

Using information on single auxiliary variable (ψ), the estimator $\hat{C}^2_{RA}$ for estimating the square of population coefficient of variation $C'^2_y = \frac{\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}{\bar{Y}^2}$ of characteristic under study y is proposed as follows;

$$\hat{C}^2_{RA} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left\{\bar{y}\left(\frac{P}{p}\right)^k\right\}^2\right]}{\bar{y}^2} \tag{1.1}$$

where k is the characterizing scalar determined by minimizing the mean square error of $\hat{C}^2_{RA}$ and $\bar{y}$ is the sample mean of y values, p and P are sample mean and population mean of auxiliary variable (ψ) respectively.

i.e. $$p = \frac{1}{n}\sum_{i=1}^{n}\psi_i \quad \text{and} \quad P = \frac{1}{N}\sum_{i=1}^{N}\psi_i$$

Assuming that first n units have been selected in the sample from N units of the population using simple random sample without replacement.

Let, $c'^2_y = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}{\bar{y}^2}$ be the usual estimator of C'^2_y

Further, let $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^r(\psi_i - P)^s$ and $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$, $y_i^2 = z_i$

$$\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i \ , \quad Y_i^2 = Z_i$$

Now, let $$\bar{y} = \bar{Y} + e_0 , \quad p = P + e_1 , \quad \bar{z} = \bar{Z} + e_2 \tag{1.2}$$

So that $E(e_0) = E(e_1) = E(e_2) = 0$

From equation (1.1), $$\hat{C}^2_{RA} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left\{\bar{y}\left(\frac{P}{p}\right)^k\right\}^2\right]}{\bar{y}^2}$$

$$= \frac{\left(1-\frac{1}{n}\right)^{-1}\left[(\bar{Z}+e_2) - \left\{(\bar{Y}+e_0)\left(\frac{P}{P+e_1}\right)^k\right\}^2\right]}{(\bar{Y}+e_0)^2}$$

$$= \frac{\left(1-\frac{1}{n}\right)^{-1}}{\bar{Y}^2}\left[\sigma_Y^2 + e_2 - 2\bar{Y}(1+C_y^2)e_0 + 2k\frac{\bar{Y}^2}{P}e_1 - (2k^2+k)\frac{\bar{Y}^2}{P^2}e_1^2 - 2\frac{e_0 e_2}{\bar{Y}} + 3(1+C_y^2)e_0^2\right] \tag{1.3}$$

up to order $\left(\frac{1}{n}\right)$.

Now taking expectation on both sides and using this value, we get,

$$Bias\left(\hat{C}^2_{RA}\right) = E\left(\hat{C}^2_{RA}\right) - C_Y^2$$

$$= \frac{C_Y^2}{n} - \left[\frac{2}{\bar{Y}^3}E(e_0 e_2) + \frac{(2k^2+k)}{P^2}E(e_1^2) - \frac{3}{\bar{Y}^2}(1+C_y^2)E(e_0^2)\right] \tag{1.4}$$

Now, using the following expression for a simple random sample of size n without replacement from a population of size N,

$$E(e_0^2)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\mu_{20},\; E(e_1^2)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\mu_{02}$$

$$E(e_0e_1)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\mu_{11} \quad E(e_1e_2)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left(\mu_{21}+2\overline{Y}\mu_{11}\right)$$

$$E(e_2^2)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left(\mu_{40}+4\overline{Y}\mu_{30}+4\overline{Y}^2\mu_{20}-\mu_{20}^2\right)$$

$$E(e_0e_2)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left(\mu_{30}+2\overline{Y}\mu_{20}\right)$$

Now from equation (1.4) we have,

$$Bias\left(\hat{C}_{RA}^2\right)=\frac{C_Y^2}{n}-\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left[\frac{2}{\overline{Y}^3}\mu_{30}+\frac{(2k^2+k)}{P^2}\mu_{02}-(3C_y^2-1)\frac{\mu_{20}}{\overline{Y}^2}\right] \tag{1.5}$$

2. Mean Square Error (MSE) Of Estimator $\hat{C}_{RA}^2$

The Mean Square error of estimator $\hat{C}_{RA}^2$, to the first order of approximation, is given by

$$MSE\left(\hat{C}_{RA}^2\right)=E\left(\hat{C}_{RA}^2-C_Y^2\right)^2$$

$$=E\left[e_2-2\overline{Y}(1+C_y^2)e_0+2k\frac{\overline{Y}^2}{P}e_1\right]^2$$

Substituting the values of expectations involved, we get

$$MSE\left(\hat{C}_{RA}^2\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left[\mu_{40}-\mu_{20}^2+4k\frac{\overline{Y}^2}{P}\mu_{21}+4k^2\frac{\overline{Y}^4}{P^2}\mu_{02}\right.$$
$$\left.+4\overline{Y}^2C_y^4\mu_{20}-8k\frac{\overline{Y}^3}{P}C_y^2\mu_{11}-4\overline{Y}C_y^2\mu_{30}\right] \tag{2.1}$$

3. Minimum MSE Of $\hat{C}_{RA}^2$

Using principle of maxima and minima we obtain optimum value of k from equation (2.1) minimizing $MSE\left(\hat{C}_{RA}^2\right)$ as follows:

$$k=-\frac{P}{2\overline{Y}^2}\frac{1}{\mu_{02}}(\mu_{21}-2\overline{Y}C_y^2\mu_{11})=k_0(say) \tag{3.1}$$

for this value of k_0, MSE of $\hat{C}_{RA}^2$ is minimum and is given by,

$$MinMSE\ (\hat{C}_{RA}^2)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left[\mu_{40}-\mu_{20}^2-4\overline{Y}C_y^2\mu_{30}+\right.$$
$$\left.4\overline{Y}^2C_y^4\mu_{20}-\frac{1}{\mu_{02}}(\mu_{21}-2\overline{Y}C_y^2\mu_{11})^2\right] \tag{3.2}$$

Also we know that

$$MSE\left(\hat{C}_y^2\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{N}{N-1}\left[\mu_{40}-\mu_{20}^2-4\overline{Y}C_y^2\mu_{30}+4\overline{Y}^2C_y^4\mu_{20}\right] \tag{3.3}$$

It may clearly be seen from equation (3.2) and (3.3) that the estimator $\hat{C}_{RA}^2$ has less mean square error than that of $\hat{C}_y^2$, and therefore, $\hat{C}_{RA}^2$ is more efficient than the conventional estimator $\hat{C}_y^2$, in the sense of having lesser MSE.

4. Estimation of Optimum Class of Estimators

In case of optimum value k_0 of k in (3.1) or its guessed value being unknown that is the population parameters involved in the expression of k_0 may be unknown in practice, the alternative is to replace these parameters by their estimates from sample values. Thus for large N or simple random sampling with replacement, replacing $\mu_{21}, \mu_{11}, \mu_{02}, C_y^2$ and $\bar{Y}$ by their estimators.

$$\hat{\mu}_{21} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2(\psi_i - p) \qquad \hat{\mu}_{11} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(\psi_i - p)$$

$$\hat{\mu}_{02} = \frac{1}{n-1}\sum_{i=1}^{n}(\psi_i - p)^2 \qquad \hat{C}'^2_y = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}{\bar{y}^2}$$

and $\hat{\bar{Y}} = \bar{y}$ respectively, we get the estimated optimum value $\hat{k}_0$ to be

$$\hat{k}_0 = -\frac{P}{2\bar{y}^2}\frac{1}{\hat{\mu}_{02}}(\hat{\mu}_{21} - 2\bar{y}c'^2_y\,\hat{\mu}_{11}) \tag{4.1}$$

Some times in practical situations μ_{02} may be known in advance as it is related to auxiliary variable ψ and in such situations we do not need $\hat{\mu}_{02}$.Hence the mean square error of the resulting estimator $\hat{C}^{2*}_{RA}$ depending on the estimated optimum value of k as $\left[\hat{k}_0\right]_{opt}$ and to the first order of approximation is given by,

$$\hat{C}^{2*}_{RA} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left\{\bar{y}\left(\frac{P}{p}\right)^{\hat{k}_0}\right\}^2\right]}{\bar{y}^2} \tag{4.2}$$

Assuming $\hat{\mu}_{21} = \mu_{21} + e_3$, $\hat{\mu}_{02} = \mu_{02} + e_4$, $\hat{\mu}_{11} = \mu_{11} + e_5$

$\hat{c}'^2_y = \hat{C}'^2_y + e_6$ with $E(e_i) = 0$, $i = 3,4,5,6$.

$$\hat{C}^{2*}_{RA} = \frac{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left\{\bar{y}\left(\frac{P}{p}\right)^{\hat{k}_0}\right\}^2\right]}{\bar{y}^2}$$

$$= \frac{\left(1-\frac{1}{n}\right)^{-1}}{\bar{Y}^2}\left[\sigma_Y^2 + e_2 - 2\bar{Y}(1 + C_y^2)e_0 + 2\hat{k}\frac{\bar{Y}^2}{P}e_1 - (2\hat{k}^2 + \hat{k})\frac{\bar{Y}^2}{P^2}e_1^2 - 2\frac{e_0 e_2}{\bar{Y}} + 3(1 + C_y^2)e_0^2\right] \tag{4.3}$$

up to order $\left(\frac{1}{n}\right)$.

Now the Mean Square error of estimator $\hat{C}^{2*}_{RA}$ is given by,

$$MSE\left(\hat{C}^{2*}_{RA}\right) = E\left(\hat{C}^{2*}_{RA} - C_Y^2\right)^2$$

$$= E\left[e_2 - 2\bar{Y}(1 + C_y^2)e_0 - \frac{M}{\mu_{02}} e_1 \right]^2 \qquad where \quad , \; M = \left\{ \mu_{21} - 2\bar{Y} C_y^2 \mu_{11} \right\}$$

$$= E(e_2^2) + 4\bar{Y}^2 (1 + C_y^2)^2 E(e_0^2) + \frac{M^2}{\mu_{02}^2} E(e_1^2) + 4\bar{Y}(1 + C_y^2)\frac{M}{\mu_{02}} E(e_0 e_1)$$

$$- 4\bar{Y}(1 + C_y^2) E(e_0 e_2) - 2\frac{M}{\mu_{02}} E(e_1 e_2)$$

Substituting values of expectation terms involved, we get

$$MSE\left(\hat{C}_{RA}^{2*}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{N}{N-1}\left[\mu_{40} - \mu_{20}^2 - 4\bar{Y}C_y^2 \mu_{30} + 4\bar{Y}^2 C_y^4 \mu_{20} - \frac{M^2}{\mu_{02}} \right]$$

which is same as $MinMSE(\hat{C}_{RA}^2)$ given by (3.2).

Thus, if the optimum value of k is not known, it can be estimated by (4.1) and the mean square error of the resulting estimator $\hat{C}_{RA}^{2*}$ is equal to that of $\hat{C}_{RA}^2$ with optimum k given by (3.2). So in the light of practical utility $\hat{C}_{RA}^{2*}$ may be preferred to $\hat{C}_{RA}^2$ without loosing efficiency.

5. Empirical Example

From the data dealing with the number of Labourers X (in thousands) and the quantity of raw materials Y (in lakes of bales) for the population size N = 20, the required value of population parameters are calculated, the variables Y and X are positively correlated. Here we associate a qualitative variable ψ with X and take $\psi = 1$ If X > 400 & $\psi = 0$ if X $\leq$ 400. Further to study the property of proposed estimator, random sample of size 8 is taken and required sample values has been calculated. [Data Source: P.Mukhopadhyay, p. 96]

$\bar{Y} = 41.5$, $P = 0.55$, $C_y^2 = 0.052808$, $\mu_{30} = 1058.5$, $\mu_{21} = -40120715$,

$\mu_{40} = 35456.4125$, $\mu_{11} = 1.825$, $\mu_{20} = 194834.208$.

And for n = 08

$p = 0.428$, $\bar{y} = 44$, $\hat{\mu}_{02} = 194941.4170$, $c_y^2 = 0.05916$, $\hat{\mu}_{30} = 2570.3393$

$\hat{\mu}_{21} = -50516.9520$, $\hat{\mu}_{40} = 71240.0625$, $\hat{\mu}_{20} = 114.5357$, $\hat{\mu}_{11} = -1101.5179$

Using above values, we get

$$MSE\left(\hat{C}_{RA}^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{N}{N-1}\right)(11387.25186)$$

and

$$MSE\left(\hat{C}_y^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{N}{N-1}\right)(19652.29795)$$

The relative efficiency of $\hat{C}_{RA}^2$ with respect to $\hat{C}_y^2$ is,

$$E = \frac{MSE\left(\hat{C}_y^2\right)}{MSE\left(\hat{C}_{RA}^2\right)} x100\ \% = 172.5816\ \%$$

Showing that the proposed estimator $\hat{C}_{RA}^2$ is better than $\hat{C}_y^2$ in the sense of having lesser mean square error. Relative efficiency of $\hat{C}_{RA}^2$ over $\hat{C}_y^2$ based on estimated MSE of $\hat{C}_{RA}^2$ and $\hat{C}_y^2$ is

$$E = \frac{M\hat{S}E\left(\hat{C}_y^2\right)}{M\hat{S}E\left(\hat{C}_{RA}^2\right)} x100\,\% = 142.5535\ \%$$

References

1. Arcos Cebrian, A. and Rueda Gracia, M. (1997). Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator, Metrika, 45, p. 171-178.

2. Changbao, WU and Sitter, R.R. (2001). Variance estimation for finite population distribution functions with complete auxiliary information, Canadian Journal of Statistics, Vol.29, No.2, p. 289-307.

3. Chien Fu Wu (1982). Estimation of variance of the ratio estimator, Biometrika, 69, 1, p. 183-189.

4. Eberhardt, Keith R. (1975). Variance estimates for the ratio estimator, Sankhya, Series C, 37, p.43-52.

5. Gupta, R.K. and Misra, S. (2006). Estimation of population variance using ratio type estimator, Indian Journal of Mathematics and Mathematical Sciences, Vol.2, No.2, p. 169-176.

6. Ray, S.K., Sahai, A. and Sahai, A. (1979). A note on ratio and product type estimators, Ann. Inst. Statist. Math., 31, parts A, p. 141-144.

7.Singh, H.P. and Singh, R. (2001). Improved ratio-type estimator for variance using auxiliary information, Jour. Ind. Soc. Ag. Statistics, 54(3), p. 276-287.

8, Srivastava, S.K. and Jhajj, H.S. (1980). A class of estimators using auxiliary information for estimating finite population variance, Sankhya, C42, p. 87-96.