Research Journal of Applied Sciences, Engineering and Technology 7(9): 1896-1899, 2014
DOI:10.19026/rjaset.7.479
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: June 28, 2013 Accepted: July 16, 2013 Published: March 05, 2014

Research Article

On Computing Mean Square Error of Ratio Estimator

[1]Muhammad Ismail, [1]Muhammad Qaiser Shahbaz, [1]Saman Shahbaz,
[2]Muhammad Rashad and [2]Muhammad Hanif
[1]Department of Statistics, COMSATS Institute of Information Technology,
[2]National College of Business, Administration and Economics, Lahore, Pakistan

Abstract: Mean Square Error (MSE) of estimators in survey sampling is generally derived by considering terms to O (n^{-1}). In this study we propose a general method to compute the mean square error of classical ration estimator up to any degree of accuracy. The method has been proposed for ratio estimator in single phase and two phase sampling. We have used the proposed method to compute mean square error to various degree of approximation. Numerical example has also been given for illustration.

Keywords: Auxiliary variable, higher order, Mean Square Error (MSE)

INTRODUCTION

Statisticians always show much concern to the variability and try to overcome this problem. They made efforts to control or minimize variability by using different statistical methods.

In estimation theory, survey statisticians have developed many estimators and compared them on the basis of their variances or mean square errors.

The scientific development in the field of survey sampling has long history but the groundbreaking work in this field is done by Neyman (1934). Cochran (1940) appears to be the first to use auxiliary information in ratio estimator. The highly correlated auxiliary variable 'X' increases the efficiency and precision of the estimators of population characteristics.

A large number of estimators have been constructed in single phase and two phase sampling by modifying regression, ratio and product estimators. The mean square errors of most of the existing estimators are derived by using Taylor's series up to first order approximation by ignoring the higher order and product terms which reduces precision of the estimators. Many survey statisticians using above sampling technique like Cochran (1977), Rao (1986), Khare and Srivastava (1993), Tripathi and Khare (1997), Tabasum and Khan (2004), Singh and Kumar (2008) and Ismail *et al.* (2011, 2012).

MATERIALS AND METHODS

The following notations will be used for deriving the higher order mean square error of classical ratio estimator:

$$\left.\begin{array}{l} \bar{y} = \bar{Y} + \bar{e}_y \,;\; \bar{y}_2 = \bar{Y} + \bar{e}_{y_2} \\ \bar{x} = \bar{X} + \bar{e}_x \,;\; \bar{x}_2 = \bar{X} + \bar{e}_{x_2} \end{array}\right\} \qquad (1)$$

We will use above transformations in deriving the mean square error of ratio estimator up to any degree of accuracy. We will also assume that joint distribution of X and Y is bivariate normal. We will use following well known result about higher joint moments of bivariate normal distribution in deriving the mean square error:

$$E\left(\bar{e}_x^{2r}\bar{e}_y^{2s}\right) = L\theta S_x^{2r}S_y^{2s} \;;\; E\left(\bar{e}_x^{2r+1}\bar{e}_y^{2s+1}\right) = M\theta S_x^{2r+1}S_y^{2s+1}$$

where, $\theta = n^{-1} - N^{-1}$ also

$$L = \mu_{2r,2s} = \frac{(2r)!\,(2s)!}{2^{r+s}} \sum_{t=0}^{p} \frac{(2\rho)^{2t}}{(r-t)!\,(s-t)!\,(2t)!} \;;\; r,\ s = 0, 1, 2, \ldots \qquad (2)$$

$$M = \mu_{2r+1,2s+1} = \frac{(2r+1)!\,(2s+1)!}{2^{r+s}}\,\rho \sum_{t=0}^{p} \frac{(2\rho)^{2t}}{(r-t)!\,(s-t)!\,(2t+1)!} \qquad (3)$$

and "p" is minimum of r and s. We also have $\mu_{r,s} = 0$ when $r + s$ is odd.

Materials: Following estimators have been proposed from time to time by different statisticians. The given mean square errors are computed up to O(n^{-1}) and ignoring the product moments.

Corresponding Author: Muhammad Ismail, Department of Statistics, COMSATS Institute of Information Technology, Lahore, Pakistan

Classical ratio estimator: The classical ratio estimator for single phase sampling is:

$$t_1 = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{4}$$

The MSE of (4) for single phase by ignoring the product term is:

$$MSE(t_{1E}) = \theta\bar{Y}^2\left[C_y^2 + C_x^2 - 2C_xC_y\rho\right]. \tag{5}$$

The ratio estimator in two phase, when population mean of X is known, is as below:

$$t_{1(2)} = \frac{\bar{y}_2}{\bar{x}_2}\bar{X} \quad (\bar{X} \text{ is known}) \tag{6}$$

The MSE of Eq. (6) classical ratio estimator for two phase by ignoring the product term is:

$$MSE(t_{1(2)E}) = \theta_2\bar{Y}^2\left[C_y^2 + C_x^2 - 2C_xC_y\rho\right]. \tag{7}$$

In the following section we have derived the mean square error of ratio estimator in single phase and two phase sampling up to any degree of accuracy.

RESULTS AND DISCUSSION

General expressions for mean square error of ratio estimator: In this section we have derived the mean square error of classical ratio estimator up to any degree of accuracy for single phase and two phase sampling.

For single phase sampling: The classical ratio estimator for single phase sampling when the mean of auxiliary variable is known, is given as:

$$t_1 = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{8}$$

Using Eq. (1) in above equation, we have:

$$t_1 = \frac{\bar{Y} + \bar{e}_y}{\bar{X} + \bar{e}_x}\bar{X}$$

Expanding and rearranging, we have:

$$t_1 - \bar{Y} = \left(\bar{Y} + \bar{e}_y\right)\sum_{j=0}^{\infty} a_j\bar{e}_x^j - \bar{Y} \tag{9}$$

where,

$$a_j = \left(\frac{-1}{\bar{X}}\right)^j$$

Squaring and applying expectation on Eq. (9), the mean square error of ratio estimator in single phase sampling is given as:

$$MSE_g(t_1) = \bar{Y}^2\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}a_ia_jE\left(\bar{e}_x^{i+j}\right) + \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}a_ia_jE\left(\bar{e}_x^{i+j}\bar{e}_y^2\right) + 2\bar{Y}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}a_ia_jE\left(\bar{e}_x^{i+j}\bar{e}_y\right)$$
$$+\bar{Y}^2 - 2\bar{Y}^2\sum_{j=0}^{\infty}a_jE\left(\bar{e}_x^j\right) - 2\bar{Y}\sum_{j=0}^{\infty}a_jE\left(\bar{e}_x^j\bar{e}_y\right).$$

Using Eq. (2) and (3), the general expression for MSE of ratio estimator in single phase sampling is:

$$MSE_g(t_1) = \bar{Y}^2\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}a_ia_jE\left(\bar{e}_x^{i+j}\right) + \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}a_ia_jE\left(\bar{e}_x^{i+j}\bar{e}_y^2\right) + 2\bar{Y}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}a_ia_jE\left(\bar{e}_x^{i+j}\bar{e}_y\right) \tag{10}$$
$$+\bar{Y}^2 - 2\bar{Y}^2\sum_{j=0}^{\infty}a_jE\left(\bar{e}_x^j\right) - 2\bar{Y}\sum_{j=0}^{\infty}a_jE\left(\bar{e}_x^j\bar{e}_y\right).$$

Expression (10) can be used to derive the mean square error of ratio estimator up to any degree of accuracy. We have derived some approximate expressions below.

Expand (10) up to $O(n^{-1})$ the Mean Square Error of ratio estimator is:

$$MSE_1(t_1) = MSE(t_{1E}) + \theta\bar{Y}^2\left(1 + 2\rho^2\right)C_x^2C_y^2.$$

The above expression is different from (5) because in (5) the product term has been ignored while deriving the mean square error. Again expanding (10) up to $O(n^{-2})$, the Mean Square Error is:

$$MSE_2(t_1) = MSE_1(t_1) + \theta\bar{Y}^2\left[3C_x^4 + 2\left(1 + 2\rho^2\right)C_x^2C_y^2 + 3\left(1 + 4\rho^2\right)C_x^4C_y^2 - 12C_x^3C_y\rho\right].$$

The mean square error up to $O(n^{-3})$ is obtained below by expanding (10) upto $O(n^{-3})$:

$$MSE_3(t_1) = MSE_2(t_1) + \theta\bar{Y}^2\begin{bmatrix}6C_x^4 + 15C_x^6 - 6C_x^3C_y\rho - 60C_x^5C_y\rho \\ +6\left(1 + 4\rho^2\right)C_x^4C_y^2 + 15\left(1 + 6\rho^2\right)C_x^6C_y^2\end{bmatrix}.$$

Expand up to $O(n^{-4})$, the Mean Square Error for $O(n^{-4})$ order is given below:

$$MSE_4(t_1) = MSE_3(t_1) + \theta\bar{Y}^2\begin{bmatrix}30C_x^6 + 105C_x^8 - 60C_x^5C_y\rho - 420C_x^7C_y\rho + 6\left(1 + 4\rho^2\right)C_x^4C_y^2 \\ +30\left(1 + 6\rho^2\right)C_x^6C_y^2 + 105\left(1 + 8\rho^2\right)C_x^8C_y^2\end{bmatrix}.$$

In similar way, Eq. (10) can be used to derive mean square error upto any order.

For two phase sampling: The classical ratio estimator for two phase sampling when the mean of auxiliary variable is known, is given as:

$$t_{1(2)} = \frac{\bar{y}_2}{\bar{x}_2}\bar{X} \tag{11}$$

Using Eq. (1) in above equation, we have:

$$t_{1(2)} = \frac{\bar{Y} + \bar{e}_{y_2}}{\bar{X} + \bar{e}_{x_2}}\bar{X}$$

Expanding and rearranging, we have:

$$t_{1(2)} - \bar{Y} = \left(\bar{Y} + \bar{e}_{y_2}\right) \sum_{j=0}^{\infty} a_j \bar{e}_{x_2}^{j} - \bar{Y} \tag{12}$$

where,

$$a_j = \left(\frac{-1}{\bar{X}}\right)^j$$

Squaring and applying expectation on Eq. (12), the mean square error of ratio estimator in two phase sampling is given as:

$$MSE\left(t_{1(2)}\right) = \bar{Y}^2 \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i a_j E\left(\bar{e}_{x_2}^{i+j}\right) + \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i a_j E\left(\bar{e}_{x_2}^{i+j}\bar{e}_{y_2}^{2}\right) + 2\bar{Y}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i a_j E\left(\bar{e}_{x_2}^{i+j}\bar{e}_{y_2}\right) + \bar{Y}^2 - 2\bar{Y}^2 \sum_{j=0}^{\infty} a_j E\left(\bar{e}_{x_2}^{j}\right) - 2\bar{Y}\sum_{j=0}^{\infty} a_j E\left(\bar{e}_{x_2}^{j}\bar{e}_{y_2}\right)$$

Using Eq. (2) and (3), the general expression for MSE of ratio estimator in two phase sampling is:

$$MSE\left(t_{1(2)}\right) = \bar{Y}^2 \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i a_j E\left(\bar{e}_{x_2}^{i+j}\right) + \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i a_j E\left(\bar{e}_{x_2}^{i+j}\bar{e}_{y_2}^{2}\right) + 2\bar{Y}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_i a_j E\left(\bar{e}_{x_2}^{i+j}\bar{e}_{y_2}\right) + \bar{Y}^2 - 2\bar{Y}^2 \sum_{j=0}^{\infty} a_j E\left(\bar{e}_{x_2}^{j}\right) - 2\bar{Y}\sum_{j=0}^{\infty} a_j E\left(\bar{e}_{x_2}^{j}\bar{e}_{y_2}\right) \tag{13}$$

Expression (13) can be used to derive the mean square error of ratio estimator up to any degree of accuracy. We have derived some approximate expressions below.

Expand (13) up to $O(n^{-1})$ the Mean Square Error of ratio estimator is:

$$MSE_1\left(t_{1(2)}\right) = MSE\left(t_{1(2)E}\right) + \theta_2 \bar{Y}^2\left(1+2\rho^2\right)C_x^2 C_y^2.$$

The above expression is different from (7) because in (7) the product term has been ignored while deriving the mean square error. Again expanding (13) up to $O(n^{-2})$, the mean square error is:

$$MSE_2\left(t_{1(2)}\right) = MSE_1\left(t_{1(2)}\right) + \theta_2\bar{Y}^2\left[3C_x^4 + 2\left(1+2\rho^2\right)C_x^2C_y^2 + 3\left(1+4\rho^2\right)C_x^4C_y^2 - 12C_x^3C_y\rho\right].$$

The mean square error upto $O(n^{-3})$ is obtained below by expanding (10) upto $O(n^{-3})$:

$$MSE_3\left(t_{1(2)}\right) = MSE_2\left(t_{1(2)}\right) + \theta_2\bar{Y}^2\left[\begin{array}{l}6C_x^4 + 15C_x^6 - 6C_x^3C_y\rho - 60C_x^5C_y\rho \\ + 6\left(1+4\rho^2\right)C_x^4C_y^2 + 15\left(1+6\rho^2\right)C_x^6C_y^2\end{array}\right].$$

Expand up to $O(n^{-4})$, the Mean Square Error for $O(n^{-4})$ order is given below:

$$MSE_4\left(t_{1(2)}\right) = MSE_3\left(t_{1(2)}\right) + \theta_2\bar{Y}^2\left[\begin{array}{l}30C_x^6 + 105C_x^8 - 60C_x^5C_y\rho - 420C_x^7C_y\rho + 6\left(1+4\rho^2\right)C_x^4C_y^2 \\ + 30\left(1+6\rho^2\right)C_x^6C_y^2 + 105\left(1+8\rho^2\right)C_x^8C_y^2\end{array}\right].$$

In similar way, (13) can be used to derive mean square error up to any order.

NUMERICAL STUDY

In this section we have given the numerical illustrations to see the effect of higher orders on mean square error of ratio estimator in single phase and two phase sampling.

We have used the data given by Khare and Sinha (2007) and Singh and Kumar (2011) for the numerical study.

The data is on physical growth of upper socioeconomic group of 95 school children of Varanasi under an ICMR study, Department of pediatrics, B.H.U., during 1983-84. We have used following two set of variables as two different populations:

Population 1:

y: Weight in kg of the children
x: Skull circumference in cm of the children

For second population we have used chest circumference in cm of the children as auxiliary variable. The results of mean square errors are given in the Table 1 and 2.

The results for two phase sampling are given in the Table 3 and 4.

Table 1: Population 1, N = 95, $\bar{Y} = 19.4968$, Cy = 0.15613, Cx = 0.03006, $\rho = 0.328$

Order	n = 10	n = 15	n = 18	n = 21	n = 25
Existing	0.755096	0.473786	0.380016	0.313037	0.248738
$O(n^{-1})$	0.756007	0.474357	0.380474	0.313414	0.249037
$O(n^{-2})$	0.757346	0.475197	0.381148	0.31397	0.249479
$O(n^{-3})$	0.757232	0.475126	0.381091	0.313923	0.249441
$O(n^{-4})$	0.757236	0.475129	0.381093	0.313924	0.249443

Table 2: Population 2, N = 95, $\bar{Y} = 19.4968$, Cy = 0.15613, Cx = 0.05860, $\rho = 0.846$

Order	n = 10	n = 15	n = 18	n = 21	n = 25
Existing	0.41936	0.263128	0.211051	0.173852	0.138142
$O(n^{-1})$	0.426283	0.267471	0.214534	0.176722	0.140422
$O(n^{-2})$	0.430596	0.270178	0.216705	0.17851	0.141843
$O(n^{-3})$	0.427642	0.268324	0.215218	0.177286	0.14087
$O(n^{-4})$	0.427725	0.268376	0.21526	0.17732	0.140898

Table 3: Population 1, N = 95, $\bar{Y} = 19.4968$, Cy = 0.15613, Cx = 0.02478, ρ = 0.477

Order	$n_1 = 6$ $n_2 = 4$	$n_1 = 10$ $n_2 = 8$	$n_1 = 15$ $n_2 = 12$	$n_1 = 21$ $n_2 = 19$	$n_1 = 25$ $n_2 = 23$
Existing	0.674713	0.202414	0.134943	0.040584	0.028162
$O(n^{-1})$	0.675402	0.202621	0.13508	0.040626	0.028191
$O(n^{-2})$	0.676389	0.202917	0.135278	0.040685	0.028232
$O(n^{-3})$	0.676247	0.202874	0.135249	0.040677	0.028226
$O(n^{-4})$	0.67625	0.202875	0.13525	0.040677	0.028226

Table 4: Population 2, N = 95, $\bar{Y} = 19.4968$, Cy = 0.15613, Cx = 0.05402, ρ = 0.729

Order	$n_1 = 6$ $n_2 = 4$	$n_1 = 10$ $n_2 = 8$	$n_1 = 15$ $n_2 = 12$	$n_1 = 21$ $n_2 = 19$	$n_1 = 25$ $n_2 = 23$
Existing	0.475085	0.142526	0.095017	0.028577	0.01983
$O(n^{-1})$	0.479734	0.14392	0.095947	0.028856	0.020024
$O(n^{-2})$	0.483081	0.144924	0.096616	0.029058	0.020163
$O(n^{-3})$	0.481326	0.144398	0.096265	0.028952	0.02009
$O(n^{-4})$	0.481374	0.144412	0.096275	0.028955	0.020092

CONCLUSION

The numerical values of mean square error of ratio estimator in single phase and two phase sampling are given in Table 1 to 4. The result shows that when higher order terms are included in calculation of mean square error the value increases from the existing one. The reason for this is positive correlation between study and auxiliary variable and inclusion of more terms in computing the mean square error.

REFERENCES

Cochran, W.G., 1940. The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. J. Agri. Sci., 30: 262-275.

Cochran, W.G., 1977. Sampling Techniques. John Wiley and Sons, New York.

Ismail, M., M.Q. Shahbaz and M. Hanif, 2011. A general class of estimator of population mean in the presence of non-response. Pak. J. Stat., 27(4): 467-476.

Ismail, M., M. Hanif and M.Q. Shahbaz, 2012. A simple generalized ratio-cum-regression estimator for population mean in the presence of non-response for two-phase sampling. Submitted to Pakistan Journal of Statistics (Seen by Courtesy of the Authors).

Khare, B.B. and R.R. Sinha, 2007. Estimation of the ratio of two population means using multi auxiliary characters of non-response. Statistical Techniques in Life Testing, Reliability, Sampling Theory and Quality Control, pp: 163-171.

Khare, B.B. and S. Srivastava, 1993. Estimation of population mean using auxiliary character in presence of non-response. Nat. Acad. Sci. Lett. India, 16: 111-114.

Neyman, J., 1934. On the two different aspects of representative method: The method of stratified sampling and the method of purposive selection. J. Roy. Stat. Soc., 97: 558-606.

Rao, P.S.R.S., 1986. Ratio estimation with sub sampling the non respondents. Surv. Methodol., 12(2): 217-230.

Singh, H.P. and S. Kumar, 2008. A regression approach to the estimation of the finite population mean in the presence of non-response. Aust. NZ. J. Stat., 50(4): 395-408.

Singh, H.P. and S. Kumar, 2011. Combination of regression and ratio estimate in presence of non response. Braz. J. Stat. Assoc., 25(2): 205-217.

Tabasum, R. and I.A. Khan, 2004. Double sampling for ratio estimation with non-response. J. Ind. Soc. Agric. Stat., 58(3): 300-306.

Tripathi, T.P. and B.B. Khare, 1997. Estimation of mean vector in the presence of non-response. Commun. Stat-Theor. M., 26(9): 2255-2269.