Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 67(3) 2013 295-299

Estimation of Population Mean using Known Coefficient of Variation

Sheela Misra, R. Karan Singh and Archana Shukla
Department of Statistics, University of Lucknow, Lucknow

Received 11 January 2012; Revised 14 November 2012; Accepted 23 November 2012

SUMMARY

A regression estimation procedure based on known coefficient of variation is proposed for the estimation of the population mean. The bias and mean square error of the proposed estimator are found. A comparative study with the usual regression estimator of the population mean has been made.

Keywords: Regression estimator, Coefficient of variation, Bias, Mean square error, Efficiency.

1. INTRODUCTION

A regression type estimator using known coefficient of variation is considered and its properties are studied. There are several instances in physical, biological and agricultural sciences where the mean is proportional to standard deviation and consequently the coefficient of variation is known although the mean and standard deviation may not be known. Some such situations may be seen in Snedecor (1946), Hald (1952), Davies and Goldsmith (1976) and Gleser and Healy (1976). The well known Weber's law of Psychophysics (see Guilford 1975 chapter 2) provides instances where coefficient of variation is known and one such example is given in Singh (1998) also.

Sometimes, simple a priori information in the form of coefficient of variation is available to the experimenters in the fields of biology, agriculture, psychometrics etc. Long association of the experimenters with the experimental material, the experimenters may have at their disposal quite accurate information concerning the coefficient of variation. This information concerning coefficient of variation is frequently used to plan experiments, estimate sample size, average, total, etc. (see Searles 1964 also). Further supporting explanation regarding stable and consistent information about coefficient of variation may be seen in Cochran (1977, 3rd edition) on page 77 and page 79 of Chapter 4. A good description about knowledge of coefficient of variation is given in Sukhatme *et al.* (1984) also on page 42.

Let the variable of interest be y and the auxiliary variable be x taking the values Y_i and X_i respectively for the i^{th} (i = 1, 2,… N) unit of the population of size N.

Further, let

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N} (X_i - \bar{X})^r (Y_i - \bar{Y})^s$$

$$C_y = \frac{\sigma_y}{\bar{Y}}, C_x = \frac{\sigma_x}{\bar{X}}$$

Corresponding Author: Sheela Misra
E-mail address: drsheelamisra@gmail.com

$$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i-\bar{Y})^2, \sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})$$

$$\beta_2 = \frac{\mu_{04}}{\mu_{02}^2}, \gamma_1 = \frac{\mu_{03}}{\mu_{02}^{3/2}}, \ \rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}, \ B = \frac{\sigma_{xy}}{\sigma_x^2}$$

Also, let

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2, \ s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}), \ b = \frac{s_{xy}}{s_x^2}$$

where $y_1, y_2, \ldots, y_n$ are the observations on y and $x_1, x_2, \ldots, x_n$ are the observations on auxiliary variable x for a simple random sample of size n.

For estimating the population mean using regression method of estimation and considering a convex combination, the proposed estimator is

$$\bar{y}_{lrk} = [\bar{y} + b(\bar{X}-\bar{x})] + k\left(\frac{s_y^2}{C_y^2} - \bar{y}^2\right) \qquad (1.1)$$

where k is the characterizing scalar chosen suitably.

2. BIAS AND MEAN SQUARE ERROR OF $\bar{y}_{lrk}$

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 term may be ignored.

Let

$$\bar{y} = \bar{Y}(1+e_0), \quad \bar{x} = \bar{X}(1+e_1),$$

$$s_y^2 = \sigma_y^2(1+e_2), \ s_x^2 = \sigma_x^2(1+e_3),$$

$$s_{xy} = \sigma_{xy}(1+e_4) \text{ so that}$$

$E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$ and

$$E(e_0^2) = \frac{\sigma_y^2}{n\bar{Y}^2} = \frac{C_y^2}{n}, \quad E(e_1^2) = \frac{\sigma_x^2}{n\bar{X}^2} = \frac{C_x^2}{n},$$

$$E(e_0e_1) = \frac{\sigma_{xy}}{n\bar{X}\bar{Y}} = \frac{\rho C_x C_y}{n}, \ E(e_2^2) = \frac{\beta_2 - 1}{n},$$

$$E(e_1e_3) = \frac{\mu_{30}}{n\bar{X}\sigma_x^2}, \quad E(e_0e_2) = \frac{\mu_{03}}{n\sigma_y^2\bar{Y}},$$

$$E(e_1e_2) = \frac{\mu_{12}}{n\sigma_y^2\bar{X}}, \quad E(e_1e_4) = \frac{\mu_{21}}{n\sigma_{xy}\bar{X}}.$$

Also, we have

$$b = \frac{s_{xy}}{s_x^2} = \frac{\sigma_{xy}(1+e_4)}{\sigma_x^2(1+e_3)}$$

$$= B(1+e_4)(1+e_3)^{-1}$$

$$= B(1+e_4)(1-e_3+e_3^2-\ldots)$$

$$= B(1-e_3+e_4+e_3^2-e_3e_4+\ldots)$$

From (1.1), writing $\bar{y}_{lrk}$ in e_i's (i = 0, 1, 2, 3, 4), we see that

$$\bar{y}_{lrk} = [\bar{Y}(1+e_0) + B(1+e_4-e_3+e_3^2-e_3e_4+\ldots)$$

$$\times\{\bar{X}-\bar{X}(1+e_1)\}] + k\left\{\frac{\sigma_y^2(1+e_2)\bar{Y}^2}{\sigma_y^2} - \bar{Y}^2(1+e_0)^2\right\}$$

$$= \bar{Y} + \bar{Y}e_0 + B(1-e_3+e_4+e_3^2-e_3e_4+\ldots)(-\bar{X}e_1)$$

$$+k\{\bar{Y}^2(1+e_2) - \bar{Y}^2(1+e_0)^2\}$$

or

$$\bar{y}_{lrk} - \bar{Y} = [\bar{Y}e_0 - B\bar{X}e_1 - B\bar{X}e_1e_4 + B\bar{X}e_1e_3 + \ldots]$$

$$+k\bar{Y}^2(e_2 - e_0^2 - 2e_0) \qquad (2.1)$$

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

$$Bias(\bar{y}_{lrk}) = E(\bar{y}_{lrk} - \bar{Y})$$

$$= \bar{Y}E(e_0) - B\bar{X}E(e_1) - B\bar{X}E(e_1e_4) + B\bar{X}E(e_1e_3)$$

$$-k\bar{Y}^2\{E(e_2) - E(e_0^2) - 2E(e_0)\}$$

$$= \rho\frac{\sigma_y\mu_{21}}{n\sigma_x\mu_{11}} + \rho\gamma_1\frac{\sigma_y}{n} - k\frac{\sigma_y^2}{n} \qquad (2.2)$$

Squaring both sides of (2.1) and taking expectation, we have mean square error of $\bar{y}_{lrk}$ up to terms of order O(1/n) to be

$$\text{MSE}\ (\bar{y}_{lrk}) = [\{\bar{Y}e_0 - B\bar{X}e_1\} + k\bar{Y}^2(e_2 - 2e_0)]^2$$

$$= \bar{Y}^2E(e_0^2) + B^2\bar{X}^2E(e_1^2) - 2B\bar{X}\bar{Y}E(e_0e_1)$$
$$+k^2\bar{y}^4\{E(e_2^2) + 4E(e_0^2) - 4E(e_0e_2)\}$$
$$+2k\bar{Y}^2\{\bar{Y}E(e_0e_2) - 2\bar{Y}E(e_0^2) - B\bar{X}E(e_1e_2)$$
$$+2B\bar{X}E(e_0e_1)\}$$

$$= \frac{\sigma_y^2}{n} + \rho^2\frac{\sigma_y^2}{n} - 2\rho^2\frac{\sigma_y^2}{n} + \frac{k^2\bar{Y}^4}{n}[(\beta_2 - 1)$$
$$+4C_y^2 - 4\gamma_1C_y] + \frac{2k\bar{Y}^3}{n}[\gamma_1C_y$$
$$-2C_y^2 - \rho\frac{\mu_{12}}{\bar{Y}\sigma_x\sigma_y} + 2\rho^2C_y^2]$$

$$= (1-\rho^2)\frac{\sigma_y^2}{n} + \frac{k^2\bar{Y}^4}{n}\{(\beta_2 - 1) + 4C_y^2 - 4\gamma_1C_y\}$$
$$+\frac{2k\bar{Y}^3}{n}\{\gamma_1C_y - 2C_y^2 - \rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}} + 2\rho^2C_y^2] \quad (2.3)$$

The optimum value of k minimizing the mean square error of $\bar{y}_{lrk}$ in (2.3) is given by

$$k = \frac{-(\gamma_1C_y - 2C_y^2 - \rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}} + 2\rho^2C_y^2)}{n[(\beta_2 - 1) + 4C_y^2 - 4\gamma_1C_y]} \quad (2.4)$$

and the minimum mean square error is given by

$$MSE\,(\bar{y}_{lrk})$$
$$= (1-\rho^2)\frac{\sigma_y^2}{n} - \frac{\bar{Y}^2[\gamma_1C_y - 2C_y^2 - \rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}} + 2\rho^2C_y^2]^2}{n[(\beta_2 - 1) + 4C_y^2 - 4\gamma_1C_y]} \quad (2.5)$$

3. ESTIMATOR BASED ON ESTIMATED OPTIMUM $\hat{k}$

If the exact or good guess of $\beta_2, \gamma_1, \rho,$ and μ_{12} are not available, we can replace these quantities by their consistent sample estimates $\hat{\beta}_2, \hat{\gamma}_1, \hat{\rho}, \hat{\mu}_{12}$ respectively and $\hat{\bar{Y}} = \bar{y}$ in (2.4) and get the estimated optimum value of k denoted by $\hat{c}$ as

$$\hat{c} = \frac{-[\hat{\gamma}_1C_y - 2C_y^2 - \hat{\rho}\frac{\hat{\mu}_{12}}{s_xs_y\bar{y}} + 2\hat{\rho}C_y^2]}{\bar{y}\,[(\hat{\beta}_2 - 1) + 4C_y^2 - 4\hat{\gamma}_1C_y]}$$

$$= \frac{-[\frac{\hat{\mu}_{03}}{s_y^3}C_y - 2C_y^2 - \hat{\rho}\frac{\hat{\mu}_{12}}{s_xs_y\bar{y}} + 2\hat{\rho}^2C_y^2]}{\bar{y}\,[\frac{\hat{\mu}_{04}}{s_y^4} - 1 + 4C_y^2 - 4\frac{\hat{\mu}_{03}}{s_y^3}C_y]} \quad (3.1)$$

where

$$\hat{\beta}_2 = \frac{\hat{\mu}_{04}}{s_y^4} \text{ with } \hat{\mu}_{04} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^4$$

$$\hat{\gamma}_1 = \frac{\hat{\mu}_{03}}{s_y^3} \text{ with } \hat{\mu}_{03} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^3$$

$$\hat{\rho} = \frac{s_{xy}}{s_xs_y},\ \hat{\mu}_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})^2$$

Thus, incorporating $\hat{c}$ in place of k in (1.1), we get the estimator based on the estimated optimum $\hat{c}$ as

$$\bar{y}_{lrc} = [\bar{y} + b(\bar{X} - \bar{x})] + \hat{c}(\frac{s_y^2}{C_y^2} - \bar{y}^2) \quad (3.2)$$

Let

$$\hat{\mu}_{03} = \mu_{03}(1+e_5), \hat{\mu}_{04} = \mu_{04}(1+e_6), \hat{\mu}_{12} = \mu_{12}(1+e_7)$$

Also, we have

$$\hat{c} = \frac{-[\text{A} - \text{B} + \text{C}]}{\bar{Y}(1+e_0)[\frac{\mu_{04}(1+e_6)}{\sigma_y^4(1+e_2)^2} + 4C_y^2 - 4\frac{\mu_{03}(1+e_5)}{\sigma_y^3(1+e_2)^{3/2}}C_y]}$$

where,

$$\text{A} = \frac{\mu_{03}(1+e_5)}{\sigma_y^3(1+e_2)^{3/2}}C_y - 2C_y^2$$

$$B = \frac{\sigma_{xy}(1+e_4)\mu_{12}(1+e_7)}{\sigma_y\sigma_x(1+e_3)^{1/2}(1+e_2)^{1/2}\sigma_x(1+e_3)^{1/2}\sigma_y(1+e_2)^{1/2}\bar{Y}(1+e_0)}$$

$$C = 2C_y^2 \frac{\sigma_{xy}^2(1+e_4)^2}{\sigma_x^2(1+e_3)\sigma_y^2(1+e_2)}$$

Therefore, $\hat{c} = \frac{-[D-E+F]}{\bar{Y}[G-H-I]}$ (3.3)

where,

$$D = \gamma_1 C_y\left(1-\frac{3}{2}e_2+e_5+...\right) - 2C_y^2$$

$$E = \rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}}(1-e_0-e_2-e_3+e_4+e_7+..)$$

$$F = 2\rho^2C_y^2(1-e_3+2e_4-e_2+..)$$

$$G = \beta_2(1-2e_2+e_0-..)$$

$$H = (1+e_0)+4C_y^2(1+e_0)$$

$$I = 4\gamma_1(1-\frac{3}{2}e_2+e_5+e_0+..)C_y$$

Writing $\bar{y}_{lrc}$ in (3.2) in terms of e_i's (i = 0, 1, 2, …, 7) and after some simplification, we have

$$\bar{y}_{lrc}-\bar{Y} = (\bar{Y}e_0 - B\bar{X}e_1) - \frac{[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}}+2\rho^2C_y^2]}{[(\beta_2-1)+4C_y^2-4\gamma_1C_y]} \times (e_2-2e_0+...) \quad (3.4)$$

Squaring both sides of (3.4), ignoring terms of e_i's greater than two and taking expectation, we have mean square error of $\bar{y}_{lrk}$ to the first degree of approximation that is, up to terms of order O(1/n) to be

$$\text{MSE}(\bar{y}_{lrc}) = E[(\bar{Y}e_0 - B\bar{X}e_1) - \frac{[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}}+2\rho^2C_y^2]}{[(\beta_2-1)+4C_y^2-4\gamma_1C_y]} \times (e_2-2e_0)]^2$$

$$= (\bar{y}^2E(e_0^2)+B^2\bar{X}^2E(e_1^2)-2B\bar{Y}\bar{X}E(e_0e_1)) + \frac{[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_y\sigma_x\bar{Y}}]^2}{[(\beta_2-1)+4C_y^2-4\gamma_1C_y]^2}[E(e_2^2)+4E(e_0^2)-4E(e_0e_2)]$$

$$-\frac{2[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_y\sigma_x\bar{Y}}]}{[(\beta_2-1)+4C_y^2-4\gamma_1C_y]}\{\bar{Y}E(e_0e_2)-2\bar{Y}E(e_0^2) - B\bar{X}E(e_1e_2)+2B\bar{X}E(e_0e_1)\}$$

$$= \frac{\sigma_y^2}{n}(1-\rho^2) - \frac{\bar{Y}^2[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_y\sigma_x\bar{Y}}+2\rho^2C_y^2]^2}{n[(\beta_2-1)+4C_y^2-4\gamma_1C_y]} \quad (3.5)$$

which shows that the estimator $\bar{y}_{lrc}$ in (3.2) based on estimated optimum $\hat{c}$ attains the same minimum mean square error of $\bar{y}_{lrk}$ in (2.6) depending on optimum value of k in (2.4).

4. CONCLUDING REMARKS

(a) From (2.5), for the optimum value of k, the estimator attains the minimum mean square error given by

$$MSE(\bar{y}_{lrk}) = (1-\rho^2)\frac{\sigma_y^2}{n} - \frac{\bar{Y}[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}}+2\rho^2C_y^2]^2}{n[(\beta_2-1)+4C_y^2-4\gamma_1C_y]} \quad (4.1)$$

(b) From (3.5), the estimator $\bar{y}_{lrc}$ depending upon estimated $\hat{c}$ optimum has the mean square error

$$MSE(\bar{y}_{lrk}) = (1-\rho^2)\frac{\sigma_y^2}{n} - \frac{\bar{Y}^2[\gamma_1C_y-2C_y^2-\rho\frac{\mu_{12}}{\sigma_x\sigma_y\bar{Y}}+2\rho^2C_y^2]^2}{n[(\beta_2-1)+4C_y^2-4\gamma_1C_y]} \quad (4.2)$$

(c) From (4.2), we see that the estimator $\bar{y}_{lrc}$ depending on estimated optimum value is always more efficient than the usual linear regression estimator $\bar{y}_{lr} = \bar{y}+b(\bar{X}-\bar{x})$ in the sense of having lesser mean square error.

(d) The use of proposed estimator is limited for the situations when coefficient of variation is known. However , in case of unknown coefficient of variation its estimated value may be used after studying the performance of the estimator (robustness) against different values of CV, if the guess is in error say 5%, 10%, 15%, 20%, 25%, 50%. Further work is being done in this direction.

5. AN ILLUSTRATION

We observe that the conditions discussed in the introduction for known coefficient of variation are satisfied for the data given in Walpole *et al.* (2005, page 473) dealing with measure of aerobic fitness is the oxygen consumption in volume per unit body weight per unit time. Thirty-one individuals were used in an experiment in order to be able to model oxygen consumption (y) against time to run one and half miles (x). Computation of required values have been done and we have the following

$\bar{Y}$ = 47.37581,

$\bar{X}$ = 10.58613,

σ_y^2 = 27.46392,

σ_x^2 = 1.86282,

μ_{04} = 2523.46629

β_2 = 3.34559,

μ_{03} = 59.71969,

C_y = 0.11062,

γ_1 = 0.41493,

ρ = – 0.86219

μ_{12} = – 2.35772, n = 31

Using the required values, we have

$MSE(\bar{y})$ = 0.88593

$MSE(\bar{y}_{lr})$ = 0.22735

$MSE(\bar{y}_{lrc})$ = $MSE(\bar{y}_{lrk})$ = 0.192428

From above, the percent relative efficiency (PRE) of the proposed estimator $\bar{y}_{lrc}$ over the mean per unit estimator $\bar{y}$ and usual linear regression estimator $\bar{y}_{lr}$ are 460% and 118% respectively, showing that the enhanced efficiency of the proposed estimator.

The Percent Relative Efficiency (PRE) of the proposed estimator over the mean per unit estimator and usual linear regression estimator

Estimators	$\bar{y}$	$\bar{y}_{lr}$
Percent Relative Efficiency	**460%**	**118%**

ACKNOWLEDGEMENTS

The present research work is financially supported by University Grant Commission (UGC) New Delhi, India under the Major Research Project grant no F. No. 40-252/2011 (SR). Authors are thankful to the referees and AE for their suggestions leading to the improvement in the paper.

REFERENCES

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

Davies, O.L. and Goldsmith, P.L. (1976). *Statistical Methods in Research and Production*. Longman Group Ltd., London.

Gleser, L.J. and Healy, J.D. (1976). Estimating the mean of a normal distribution with known coefficient of variation, *Jour. Amer. Statist. Assoc.,* **71**, 977-981.

Guilford, J.P. (1975). *Psychometric Methods.* TATA McGraw Hill Publishing Company Ltd., Bombay.

Hald, A. (1952). *Statistical Theory with Engineering Applications*. John Wiley and Sons, Inc, New York.

Searls, D.T. (1964). The utilization of known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.,* **59**, 1225-1226.

Singh, Ramkaran (1998). Sequential estimation of the mean of normal population with known coefficient of variation. *METRON,* **LVI(3-4)**, 73-90.

Snedecor, G.W. (1946). *Statistical Methods*. The Iowa State College Press, Ames, I.A.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok C. (1984). *Sampling Theory of Survey with Applications.* IOWA State University Press, Ames, Iowa (U.S.A) and Indian Society of Agricultural Statistics, New Delhi-110012 (India).

Walpole, R.E., Mayer, R.H., Mayer, S.L. and Ye, K. (2005). *Probability and Statistics for Engineers and Scientists.* Pearson Education (Singapore) Pte. Ltd., India Branch, 482 F.I.E. Patparganj, Delhi-110092, (India)