Jour. Ind. Soc. Ag. Statistics
Vol. XLII, No. 1 (1990), pp. 131-139

EFFICIENT ESTIMATORS OF MEAN OF FINITE POPULATIONS WITH KNOWN COEFFICIENT OF VARIATION

K. K. S. BISHT
V. P. K. A., Almora
and
B. V. S. SISODIA
N. D. U. A. & T., Kumarganj, Faizabad

(Received : September, 1988)

SUMMARY

Prior knowledge of coefficient of variation of study variate and information on auxiliary character are used in developing some efficient estimators of mean of finite population. It is found that the proposed estimators are more efficient than simple mean, ratio, regression and product estimator in mos tof the situations.

Keywords : Auxiliary information, Coefficient of variation.

1. Introduction

Use of information on auxiliary character x highly correlated with the character under study y to improve the estimate of mean $\bar{Y}$ of a finite population is well recognised in literature as well as in practice. Some of the important examples are ratio and regression methods of estimation. Estimators utilizing population coefficient of variation (C.V.) of y have been proposed by Searle [3], Khan [2], Hirano [1] and others. Sisodia and Dwivedi [4] modified the ratio estimator using C.V. of auxiliary character. These estimators are no doubt more efficient than the simple mean but the gain in efficiency is hardly appreciable. Assumption about prior knowledge of C.V. of study variable is not uncommon in literatrue. An investigator can obtain approximately the value of C.V. of study variable through his experience in repeated surveys. In this paper, some estimators of population mean are proposed making use of auxiliary information alongwith the knowledge of C.V. of the character under study. Expected values and variances of these estimators are deduced. The estimators, under optimum conditions, are compared with simple mean, ratio, regression and product estimators. Further practicable values for the optimum choices are suggested and the estimators thus obtained are also compared.

2. Notations

Consider that a sample of size n is drawn by simple random sampling without replacement from a population that consists of N identifiable units. Let $\bar{y}$, C_{sy} and C_y respectively denote the sample mean, sample C.V. and population C.V. for the character under study; y and C_{sx} and C_x denote the sample and population C.V. of the auxiliary character x. Let $\bar{Y}$ and $\bar{X}$ denote the population mean of the characters y and x, respectively. Let s_x^2 and s_y^2 denote sample mean square error of x and y, respectively. Similary S_x^2 and S_y^2 denote the population mean square error of x and y. Let ρ be the coefficient of correlation between y and x, $\theta = (1/n - 1/N)$ and $\mu_{rs} = [(x - \bar{X})^r (y - \bar{Y})^s]$. Assume that information on auxiliary character x on each unit of the population is available and hence C_x can be determined. It is assumed that a prior knowledge of C_y is available.

3. Proposed Estimators

The estimators of population mean $\bar{Y}$ are proposed utilising C_{sx}, C_{sy}, and C_y.

3.1 *Estimator 1*

The first estimator of $\bar{Y}$ proposed is

$$t_1 = \bar{y} + K_1 \left(\frac{C_{sx}}{C_x} - \frac{C_y}{C_{sy}} \right) \tag{1}$$

where K_1 is a scalar quantity. It may be noted that this estimator is consistent. Assume that

$$\bar{y} = \bar{Y} + \eta_1$$

$$s_y^2 = S_y^2 + \eta_2$$

$$x = \bar{X} + \varepsilon_1$$

$$s_x^2 = S_x^2 + \varepsilon_2$$

where $E(\eta_1) = E(\eta_2) = E(\epsilon_1) = E(\epsilon_2) = 0$

Under the above assumptions, the estimator t_1 can be expressed as

$$t_1 = \bar{Y} + \eta_1 + K_1\left[\left(1 + \frac{\epsilon_2}{S_x^2}\right)^{\frac{1}{2}}\left(1 + \frac{\epsilon_1}{\bar{X}}\right)^{-1} - \left(1 + \frac{\eta_1}{\bar{X}}\right)\left(1 + \frac{\eta_2}{S_y^2}\right)^{-1/2}\right]$$

We now assume that $|\epsilon_2/S_x^2|$, $\epsilon_1/X|$ and $|\eta_2/S_y^2|$ are each less than one so that their expansion is valid. Using the results of Sukhatme and Sukhatme ([5], p. 194 and appendix II on p. 190) and neglecting the terms of order $\frac{1}{n^\nu}$ where $\nu > 1$ (i.e. upto first order of approximation), the expected value of t_1 can easily be obtained as follows :

$$E(t_1) = \bar{Y} + K_1\theta \left[\frac{\mu_{20}}{\bar{X}^2} - \frac{1}{2}\left(\frac{\mu_{30}}{\mu_{20}\bar{X}} - \frac{\mu_{03}}{\mu_{02}\bar{Y}}\right)\right] \tag{2}$$

Thus, the bias in t_1 is

$$B(t_1) = K_1\theta \left[\frac{\mu_{20}}{\bar{X}^2} - \frac{1}{2}\left(\frac{\mu_{30}}{\mu_{20}\bar{X}} - \frac{\mu_{03}}{\mu_{02}\bar{Y}}\right)\right] \tag{3}$$

Under the aforesaid assumptions and the results of Sukhatme and Sukhatme [5], the mean-square-error (MSE) of t_1 upto $0(n^{-1})$ is

$$\text{MSE}(t_1) = \theta\left[S_y^2 + K_1^2\left(C_y^2 + C_x^2 + 2\rho C_x C_y\right) - 2K_1\bar{Y}\left(C_y^2 + \rho C_x C_y\right)\right] \tag{4}$$

Minimisation of MSE (t_1) with respect to K_1 gives the optimum $K_1 = K_{0(1)}$ (say) as

$$K_{0(1)} = \frac{\bar{Y}\left(C_y^2 + \rho C_x C_y\right)}{\left(C_y^2 + C_x^2 + 2\rho C_x C_y\right)} \tag{5}$$

and the minimum value of MSE (t_1) = MSE $(t_1)_0$ (say) is

$$\text{MSE } (t_1)_0 = \frac{\theta A^2}{1 + A^2 + 2A\rho} S_y^2 (1 - \rho^2); A = \frac{C_x}{C_y} \tag{6}$$

This shows clearly that MSE $(t_1)_0$ will always be less than that of simple mean $\bar{y}$. It is interesting to notice that this is true even if the characters are uncorrelated, i.e. $\rho = 0$ and moreover, there will be a 100% relative gain in efficiency over simple mean in such a case if $A = 1$ and A^{-2} percent if $A \neq 1$. This is so because known C.V. of y along with the C.V. of x is utilised in the estimator, which brings higher precision even if $\rho = 0$.

Further, as $\theta\ S_y^2\ (1 - \rho^2)$ is the MSE of regression estimator and $A^2/(1 + A^2 + 2A\ \rho)$ will always be less than unity for $0 < \rho < 1$, t_1 will always, under optimum condition (5), be more efficient than regression estimator and hence than ratio estimator.

Choice of $K_{0(1)}$ and deduced estimator

Since it is difficult to achieve the optimum value $K_{0(1)}$ in practice because $\bar{Y}$ and ρ are not known, we replace C_y by C_x and $\bar{Y}$ by $\bar{y}$ in $K_{0(1)}$ and get

$$K_{0(1)} = \frac{\bar{y}}{2} \tag{7}$$

Now, the proposed estimator with $K_1 = \bar{y}/2$ will be investigated.

Let us call this estimator t_1'. Thus

$$t_1' = \bar{y} + \frac{\bar{y}}{2}\left(\frac{C_{sx}}{C_x} - \frac{C_y}{C_{sy}}\right) \tag{8}$$

The bias and MSE t_1' to 0 (n^{-1}) are

$$B(t_1') = \theta\ \frac{\bar{Y}}{2}\left[\left(\frac{\mu_{20}}{\bar{X}^2} - \frac{\mu_{11}}{\bar{X}\bar{Y}} - \frac{\mu_{02}}{\bar{Y}^2}\right)\right.$$

$$\left. - \frac{1}{2}\left(\frac{\mu_{30}}{\mu_{20}\bar{X}} - \frac{2\mu_{03}}{\mu_{02}\bar{Y}} - \frac{\mu_{21}}{\mu_{20}\bar{Y}}\right)\right] \tag{9}$$

and

$$\text{MSE}(t_1')\,\theta = \frac{\bar{Y}^2}{4}\left[C_y^2 + C_x^2 - 2\,\rho C_x C_y\right]$$

$$= \frac{\theta\ \bar{Y}^2 C_y^2}{4}\left[1 + A^2 - 2\rho A\right] \tag{10}$$

Comparing t_1' with simple mean we have

$$3 - A^2 + 2\,\rho\,A > 0$$

$$\text{Or}\quad \rho > \frac{A^2 - 3}{2A} \tag{11}$$

for t_1' to be more efficient than simple mean. The above condition on ρ is quite feasible in the practice. The estimator t_1' will always be more efficient than ratio estimator as it is evident from (10) that MSE of ratio estimate is four times the MSE of t_1'.

On comparing regression estimator with t_1', we find that the latter is more efficient than the former if

$$A^2 - 2\,\rho\,A - 3 + 4\,\rho^2 < 0 \tag{12}$$

i.e. ρ lies in the interval $[A \pm \sqrt{3\,(4 - A^2)}]/4$.

The estimator t_1' will be more efficient than product estimator if

$$3A^2 + 10\,\rho\,A + 3 > 0$$

$$\text{i.e. } \rho > -\frac{3\,(1 + A^2)}{10A} \tag{13}$$

It is worthwhile to note that the above inequality remains unaltered when A is changed to $\frac{1}{A}$, i.e. the inequality is symmetric in C_x and C_y.

Special case : If we consider $A = 1$, i.e. $C_x = C_y$ as may be expected, for example, when y and x denote values in two consecutive periods for the same character, the inequality (11) reduces to $\rho > -1$ indicating thereby t_1' to be more efficient than simple mean whatever be the value of ρ except when $\rho = -1$. Similarly, ρ should lie between -0.5 and 1 for t_1' to be more efficient than regression estimator and, $\rho > -0.6$ for it to be more efficient than product estimator.

3.2 *Estimator* 2

The second estimator of $\bar{Y}$ is proposed as

$$t_2 = \bar{y} + K_2 \left(\frac{C_{sx}}{C_x} - \frac{C_{sy}}{C_y} \right) \tag{14}$$

where K_2 is scalar quantity. Upto $0\,(n^{-1})$, the bias and MSE are obtained as

$$B\,(t_2) = K_2\,\theta \left[\frac{\mu_{20}}{\bar{X}^2} - \frac{\mu_{02}}{\bar{Y}^2} - \frac{1}{2}\left(\frac{\mu_{30}}{\mu_{20}\,\bar{X}} - \frac{\mu_{03}}{\mu_{02}\,\bar{Y}} \right) \right] \tag{15}$$

$$\text{MSE}\;(t_2) = \theta\,[S_y^2 + K_2^2\,(C_y^2 + C_x^2 - 2\,\rho\,C_x\,C_y) + 2K_2\,\bar{Y}\,(C_y^2 - \rho\,C_x\,C_y)] \tag{16}$$

Minimising MSE (t_2) with respect to K_2, we get the optimum $K_2 = K_{0(2)}$ (say) as

$$K_{0(2)} = -\frac{\bar{Y}(C_y^2 - \rho\,C_x\,C_y)}{C_y^2 + C_x^2 - 2\rho C_x\,C_y} \tag{17}$$

and the minimum MSE as

$$\text{MSE}\,(t_2)_0 = \theta\;\;\bar{Y}^2 \left[C_y^2 - \frac{(C_y^2 - \rho\,C_x\,C_y)^2}{C_y^2 + C_x^2 - 2\,\rho\,C_x\,C_y} \right] \tag{18}$$

This shows that MSE $(t_2)_0$ will always be less than the variance of simple mean $\bar{y}$. This is true even if the characters are uncorrelated, i.e. $\rho = 0$. Moreover, there will be a 100 percent gain in efficiency if $C_x = C_y$. Further, if $C_x = A\;C_y$, the gain in efficiency is proportional to $1/A^2$. MSE $(t_2)_0$ can further be simplified as

$$\text{MSE}\,(t_2)_0 = \frac{\theta\;A^2}{1 + A^2 - 2\rho A}\;S_y^2\,(1 - \rho^2) \tag{19}$$

Since $\theta\cdot S_y^2\,(1 - \rho^2)$ is MSE of regression estimator and $A^2/(1 + A^2 - 2\rho A)$ will always be less than unity for $-1 < \rho < 0$, the proposed estimator will always be, under optimum condition (17), more efficient than regression estimator and hence than product estimator.

Choice of $K_{0(2)}$ and deduced estimator

In this case replace C_y by C_x and $\bar{Y}$ by $\bar{y}$ in $K_{0(2)}$, and the choice for $K_{0(2)}$ is

$$K_{0(2)} = -\frac{\bar{y}}{2} \tag{20}$$

Thus we get the estimator

$$t_2' = \bar{y} - \frac{\bar{y}}{2}\left(\frac{C_{sx}}{C_x} - \frac{C_{sy}}{C_y}\right) \tag{21}$$

To 0 (n^{-1}), the bias and MSE of t_2' are

$$B(t_2') = -\frac{1}{2}\,\theta\,\bar{Y}\left[\frac{\mu_{20}}{\bar{X}^2} - \frac{\mu_{11}}{\bar{X}\,\bar{Y}} + \frac{1}{2}\left(\frac{\mu_{21}}{\mu_{20}\bar{Y}} - \frac{\mu_{30}}{\mu_{20}\bar{X}}\right)\right] \tag{22}$$

and

$$\text{MSE}(t_2') = \theta\,\frac{\bar{Y}^2}{4}\left[C_y^2 + C_x^2 + 2\rho\,C_x\,C_y\right] \tag{2.3}$$

We get the following conditions on correlation coefficient ρ for t_2' to be more efficient as compared to the standard estimators.

No condition is required for t_2' to be more efficient than product estimator as MSE of product estimator is four times the MSE of t_2'. The estimator t_2' will be more efficient than simple mean, ratio estimator and regression estimator, respectively, if

$$\rho < \frac{3 - A^2}{2A}, \tag{24}$$

$$\rho < \frac{3(1 + A^2)}{10A} \tag{25}$$

and ρ lies in the interval $[-A \pm \sqrt{3}\,(4 - A)^2]/4$ (26)

Remarks :— Basically, the estimator t_1 (or t_1') is proposed for the situation where x is positively correlated with y while the estimator t_2 (or t_2')

is proposed for the situation where x is negatively correlated with y. It is very obvious from the results (see eq. 10 and 23) that no condition is required for t_1' and t_2' to be more efficient than ratio estimator and product estimator, respectively. However. both the estimators t_1' and t_2' can be used in both the situations under certain conditions and can produce result more efficient than regression estimator.

Now, if we examine the above listed conditions, then it is seen that:

(i) No condition is required for t_2' to be more efficient than product estimator and its use will lead to 300% gain in efficiency.

(ii) For $A > 1$, t_2' will be more efficient than simple mean for all value of ρ. For $A < 1$. the upper limit of ρ decreases but, however, there is still a wide range of ρ for which t_2' will be more efficient than simple mean.

(iii) t_2' will also be more efficient than regression estimator in a wide range of ρ. The range of ρ, however, decreases as A increases.

The efficiency of t_1' can also be similarly examined for different values of A. It could easily be seen that the estimators t_1' and t_2' together cover almost entire range of ρ in which at least one of them will be more efficient than simple mean, ratio, regression and product estimators and gain in efficiency is substantial. However, some specific recommendations would be as follows :

(i) In general, t_1' and t_2' should always be preferred to ratio and product estimators, respectively.

(ii) If A lies between 0.5 and 1.5, then t_1' and t_2' should always be preferred, in general, to even regression estimator for $0 < \rho < \cdot 95$ and $-0.95 < \rho < 0$ respectively.

ACKNOWLEDGEMENT

Authors are very much thankful to the referee for his valuable suggestions which brought considerable improvement in this paper.

REFERENCES

[1] Hirano, K. (1972) : Using some approximately known coefficient of variation in estimating the mean. Research memorandum No. 49, Inst. Stat. Math., Tokyo. Japan.

[2] Khan, R.A. (1968) : A note on estimation of the mean of a normal distribution with known coefficient of variation. *J. Amer. Statist. Assoc.*, **63** : 1039 – 41.

[3] Searle, D.T. (1964) : Utilisation of a known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.* 59 : 1225 — 26.

[4] Sisodia, B. V. S. and Dwivedi, V. K. (1981) : A modified ratio estimator using coefficient of variation of auxillary variable. *J. Indian. Soc. Agri. Statist.* 33 (1) : 13 — 18.

[5] Sukhatme, P. V. and Sukhatme, B. V. (1970) : *Sampling Theory of Surveys with Applications.* Indian Soc. of Agricultural Statistics, New Delhi.