Jour. Ind. Soc. Ag. Statistics
Vol. XXXVIII, No. 2 (1986), pp. 230-236

COMPARISON OF SOME ESTIMATORS OF *K* AND *B* IN FINITE POPULATIONS

S. K. SRIVASTAVA, H. S. JHAJJ AND M. K. SHARMA
Department of Statistics, Punjabi University, Patiala

(Received : June, 1985)

SUMMARY

Three estimators of $K = \rho C_y/C_x$ and two estimators of $B = S_{yx}/S_x^2$ in finite populations, are compared.

Keywords : Auxiliary information; Ratio-type estimators; Regression coefficient; Simple random sampling.

Introduction

For estimating the mean $\bar{Y}$ of a finite population using information on an auxiliary variable x, ratio-type estimators have been very widely used. Generalisation of ratio-type estimators has received considerable attention during the last fifteen years. Srivastava [6] was the first to generalise the ratio and product estimator for estimating the mean $\bar{Y}$. He considered the estimator

$$\tilde{y}_s = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)^{\alpha} \qquad (1)$$

Where the optimum value of α is $-\rho C_y/C_x = -K$, say, which was obtained by minimising the mean squared error (MSE) of $\tilde{y}_s$ up to terms of order n^{-1}. Here $\bar{y}$ and $\bar{x}$ denote the simple random sample means of the variables y and x respectively, and $\bar{X}$ denotes the known population mean of the variable x; ρ denotes the correlation coefficient between y and x, and C_y and C_x denote the coefficients of variation of y and x respectively. Several other estimators of $\bar{Y}$ have since been considered in literature in which the function $(\bar{x}/\bar{X})^{\alpha}$ in (1) has been replaced by other functions of $\bar{x}/\bar{X}$. To name a few, the following three estimators were studied by Chakrabarty [1], Walsh [11] and Vos [10].

$$\tilde{\bar{y}}_C = (1 - w)\,\bar{y} + w\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{2}$$

$$\tilde{\bar{y}}_W = \bar{y}\,\frac{\bar{X}}{A\bar{x} + (1 - A)\,\bar{X}} \tag{3}$$

$$\tilde{\bar{y}}_V = (1 - a)\,\bar{y} + a\bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) \tag{4}$$

The values of the constants w, A and a in (2), (3) and (4) were determined so as to minimise the MSE (up to terms of order n^{-1}) of the respective estimators and these optimum values are $w = K$, $A = K$ and $a = -K$.

Srivastava [7] defined a large class of estimators of $\bar{Y}$ of which the previously defined estimators are members. For the case of a single auxiliary variable, the class considered was

$$\tilde{\bar{y}}_h = \bar{y}h\left(\frac{\bar{x}}{\bar{X}}\right)$$

where $h(\cdot)$ is a parametric function such that $h(1) = 1$ and satisfies certain regularity condition. He obtained the optimum values of the parameters in $h(\cdot)$ which minimise the MSE of the estimator up to terms of order n^{-1}. The optimum values of the parameters so obtained were dependent upon K only.

All such estimators, thus, involved K which is a function of population parameters. When the value of K is not known, it is required to be estimated from the given sample. Reddy [5] has shown that the value of K is fairly stable in repeated surveys. He established that the value of K is more stable than other population parameters such as the linear regression coefficient B.

In this paper three estimators of K and two estimators of B, are compared.

2. Notations

Assume that a simple random sample of size n is drawn from the given population of size N. Let the values of the variables y and x be denoted by Y_i and X_i respectively for the ith unit of the population, $i = 1, \ldots,$ N, and by y and x for the ith unit in the sample, $i = 1, \ldots, n$. Write

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

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

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

$$\mu_{rs} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^r (X_i - \bar{X})^s,\ \lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\ \mu_{02}^{s/2}}$$

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{20}}{\bar{Y}^2},\ C_x^2 = \frac{S_x^2}{\bar{X}^2} = \frac{\mu_{02}}{\bar{X}^2},\ \rho = \frac{S_{yx}}{S_y S_x} = \frac{\mu_{11}}{S_y S_x}.$$

For simplicity assume that population size N is large as compared to the sample size n so that the finite population correction terms are ignored throughout. Writing

$$\epsilon_0 = \frac{\bar{y}}{\bar{Y}} - 1,\quad \epsilon = \frac{\bar{x}}{\bar{X}} - 1,$$

$$\delta = \frac{s_x^2}{S_x^2} - 1,\quad \eta = \frac{s_{xy}}{S_{xy}} - 1,$$

we have

$$E(\epsilon_0) = E(\epsilon) = E(\delta) = E(\eta) = 0,$$

$$E(\epsilon_0^2) = \frac{1}{n} C_y^2,\ E(\epsilon^2) = \frac{1}{n} C_x^2,$$

$$E(\epsilon_0\epsilon) = \frac{1}{n} \rho C_y C_x,\ E(\epsilon_0\,\delta) = \frac{1}{n} C_x \lambda_{12},$$

$$E(\epsilon_0\eta) = \frac{1}{n} C_y \lambda_{21}/\rho,\ E(\epsilon\delta) = \frac{1}{n} C_x \lambda_{03},$$

$$E(\epsilon\eta) = \frac{1}{n} C_x \lambda_{12}/\rho,$$

and up to terms of order n^{-1},

$$E(\delta^2) = \frac{1}{n} (\lambda_{04} - 1),\ E(\eta^2) = \frac{1}{n} \left(\frac{\lambda_{22}}{\rho^2} - 1 \right),$$

$$E(\delta\eta) = \frac{1}{n} \left(\frac{\lambda_{13}}{\rho} - 1 \right).$$

3. Estimation of K

K can be put in the form

$$K = \frac{\bar{X}}{\bar{Y}} \frac{S_{xy}}{S_x^2} .$$

Since we are considering the situation in which information on the auxiliary variable x is available, the following three estimators of K naturally come to mind

$$k_1 = \frac{\bar{x}}{\bar{y}} \frac{s_{xy}}{s_x^2}$$

$$k_2 = \frac{\bar{X}}{\bar{y}} \frac{s_{xy}}{s_x^2}$$

$$k_3 = \frac{\bar{X}}{\bar{y}} \frac{s_{xy}}{S_x^2} .$$

It is easily shown that all three estimators are consistent. The estimator k_1 does not use any knowledge of the population parameters of the auxiliary variable x, k_2 uses the knowledge of the population mean $\bar{X}$ and k_3 uses the knowledge of both mean $\bar{X}$ and variance S_x^2 of the auxiliary variable x.

To find the mean square error of k_1 substitute of $\bar{y}$, $\bar{x}$, s_x^2 and s_{xy} in it in terms of ϵ_0, ϵ, δ and η, to obtain

$$k_1 = \frac{\bar{X} S_{xy}}{\bar{Y} S_x^2} (1 + \epsilon)(1 + \eta)(1 + \epsilon_0)^{-1}(1 + \delta)^{-1}$$

$$= K(1 + \epsilon + \eta - \epsilon_0 - \delta - \epsilon_0\epsilon - \epsilon_0\eta + \epsilon_0\delta + \epsilon\eta - \epsilon\delta + \epsilon_0^2 + \delta^2 - \delta\eta + \ldots). \quad (5)$$

Taking expectation in (5) and using results from Section 2 it is easily seen that

$$E(k_1) = K + 0(n^{-1})$$

and the MSE up to terms of order n^{-1} is given by

$$M(k_1) = E(k_1 - K)^2$$

$$= \frac{K^2}{n} (C_y^2 + C_x^2 - 2\,\rho C_y C_x - 2\,\lambda_{03} C_x + 2\,\lambda_{12} C_y + \lambda_{04} + \frac{\lambda_{22}}{\rho^2} + \frac{2\lambda_{12}}{\rho} C_x - \frac{2\lambda_{21}}{\rho} C_y - \frac{2\lambda_{13}}{\rho} \quad (6)$$

Similarly it is easily found that the bias of k_2 and k_3 are of order n^{-1}, and their mean squared errors up to order n^{-1} are given by

$$M(k_2) = \frac{K^2}{n}\left(C_y^2 + 2\lambda_{12}C_y + \lambda_{04} + \frac{\lambda_{22}}{\rho^2} - \frac{2\lambda_{21}}{\rho} C_y - \frac{2\lambda_{13}}{\rho}\right) \quad (7)$$

$$M(k_3) = \frac{K^2}{n}\left(C_y^2 - 1 + \frac{\lambda_{22}}{\rho^2} - \frac{2\lambda_{21}}{\rho} C_y\right). \quad (8)$$

3.1 *Comparison of MSE*

From (6), (7) and (8) we obtain

$$M(k_1) < M(k_2) \text{ if } \rho\frac{C_y}{C_x} > \frac{1}{2} + \frac{\lambda_{12}}{\rho C_x} - \frac{\lambda_{03}}{C_x}, \quad (9)$$

$$M(k_1) < M(k_3) \text{ if } \rho\frac{C_y}{C_x} > \frac{1}{2} - \frac{\lambda_{03}}{C_x} + \frac{\lambda_{12}C_y}{C_x^2} + \frac{\lambda_{04}}{2C_x^2} + \frac{\lambda_{12}}{\rho C_x} - \frac{\lambda_{13}}{\rho C_x^2} + \frac{1}{2C_x^2}. \quad (10)$$

$$M(k_2) < M(k_3) \text{ if } \lambda_{04} + 1 + 2\lambda_{12}C_y < \frac{2\lambda_{13}}{\rho}. \quad (11)$$

In the case of bivariate normal population the inequality (9) reduces to $\rho C_y/C_x > \frac{1}{2}$, the inequality (10) reduces to $\rho C_y/C_x > \frac{1}{2} - 1/C_x^2$ and the inequality (11) is always true.

3.2. *Numerical Illustration*

For the purpose of numerical illustration of the MSEs of the three estimators k_1, k_2 and k_3, computations have been made for the five populations described in Table 3.1, which have been used in literature for illustration of the efficiency of different type of estimators using auxiliary information.

In Table 3.2 the MSE of the three estimators k_1, k_2 and k_3 up to terms of order n^{-1} are given. It is found that excepting the first population where $M(k_1) > M(k_2)$, for all other populations $M(k_1) < M(k_2) < M(k_3)$. For population I, the value of the correlation coefficient is small as compared to those of other populations.

4. Estimation of B

As in section 2, consider the two estimators of the population regression coefficient $B = S_{yx}/S_x^2$, given by

$$b_1 = \frac{s_{xy}}{s_x^2}$$

TABLE 3.1—DESCRIPTION OF POPULATIONS

Sr. No.	*Source*	*y*	*x*	ρ	C_y	C_x
1.	Cochran (1977) p. 325	No. of persons per block.	No. of room per block.	.6515	.1449	.1281
2.	Horvitz and Thompson (1952)	No. of house-hold	Eye Estimate of *y*	.8662	.4264	.3889
3.	Cochran (1977) p. 203.	Actual wt. of peaches on each tree.	Eye estimate of wt. of peaches on each tree.	.9737	.1840	.1621
4.	Sukhatme and Sukhatme (1970) p 135 vill 1-10	Wheat acreage in 1937.	Wheat acreage in 1936.	.9770	.6164	.5625
5.	Murthy (1967) p. 399 vill 1-10	Area under wheat in 1964.	Area under wheat in 1963.	.9773	.6187	.5664

TABLE 3.2—MSE OF k_1, k_2 AND k_3 UP TO TERMS OF ORDER n^{-1}

Population No.	*1/n × MSE of*		
	k_1	k_2	k_3
1.	.7245	.6743	1.3333
2.	.3026	.5194	1.3053
3.	.0839	.1926	1.4583
4.	.0427	.3542	1.3991
5.	.0411	.3530	1.4054

and

$$b_2 = \frac{S_{xy}}{S_x^2} \cdot$$

Proceeding as in section 2 obtain the MSEs of b_1 and b_2 given by

$$M(b_1) = \frac{B^2}{n}\left(\lambda_{04} - \frac{2\lambda_{13}}{\rho} + \frac{\lambda_{22}}{\rho^2}\right) \tag{12}$$

$$M(b_2) = \frac{B^2}{n}\left(\frac{\lambda_{22}}{\rho^2} - 1\right) \cdot \tag{13}$$

From (12) and (13), we have

$$M(b_1) < M(b_2) \text{ if } \lambda_{04} < \frac{2\lambda_{13}}{\rho} - 1 \tag{14}$$

The inequality (14) is always true for the case of bivariate normal population.

In Table 4.1 the MSE of b_1 and b_2 are given up to terms of order n^{-1} for the five populations described in Table 3.1. It is found that b_1 has a smaller MSE than that of b_2 for all the five populations.

TABLE 4.1—MSE OF b_1 AND b_2 UP TO TERMS OF ORDER n^{-1}

Population No.	*1/n × MSE of*	
	b_1	b_2
1.	2.0191	4.2318
2.	.6058	2.0098
3.	.0786	1.3687
4.	.0942	1.8996
5.	.0921	1.9256

REFERENCES

[1] Chakrabarty, R. P. (1968) : Contributions to the theory of ratio-type estimators, *Ph.D. Thesis*, Texas *A* and *M* University.

[2] Cochran, W. G. (1977) : *Sampling Techniques* (3rd Edition), John Wiley and Sons., N. Y.

[3] Horvitz, D. G. and Thompson, D. J. (1952) : A generalization of sampling without replacement from a finite universe. *Jour. Amer. Skt. Assoc.*, **47** : 663-685.

[4] Murthy, M. N. (1967) : *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

[5] Reddy, V. N. (1978) : A study on the use of prior knowledge on certain population parameters in estimation, *Sankhya* Ser. C, 4, 29-37.

[6] Srivastava, S. K. (1967) : An estimator using auxiliary information in sample surveys, *Cal. Stat. Assoc. Bull.*, **16** : 121-132.

[7] Srivastava, S. K. (1971) : A generalized estimator for the mean of a finite population using multi-auxiliary information, *Jour. Amer. Stat. Assoc*, **66** : 404-407.

[8] Srivastava, S. K. (1980) : A class of estimator using auxiliary information in sample survey, *Canad. Jour. Stat.*, **8** : 252-254.

[9] Sukhatme, P. V. and Sukhatme, B. V. (1970) : *Sampling Theory of Surveys with Applications* (2nd Edition), Asia Publishing House.

[10] Vos. J. W. E. (1980) : Mixing of direct, ratio and product method estimators, *Statistica Neurlandica*, **34** : 209-213.

[11] Walsh, J. N. (1970) : Generalization of ratio estimate for population total, *Sankhya* Ser. A, **32** : 99-106.