Jour. Ind. Soc. Ag. Statistics
47(2), 1995 : 119–128

Classes of Estimators of Finite Population Mean and Variance Using Auxiliary Information

Surendra K. Srivastava and Harbans Singh Jhajj
Punjabi University, Patiala
(Received: May, 1988)

SUMMARY

Srivastava and Jhajj [9], [10] have considered classes of estimators for the population mean and population variance of the study variable for a finite population using information of the population mean and variance of an auxiliary variable. In sample surveys we have information on an auxilliary variable, utilisation of which increases efficiency of estimators. In this paper classes of estimators of popuulation mean and population variance of the study variable are considered using information on the third moment of the axuxiliary variable along with its mean and variance. Asymptotic expressions for the mean square errors of the estimators in the classes and lower bounds for them are obtained.

Key words : Finite population; Auxiliary information; Asymptotic mean square error; Efficiency.

Introduction

The probolem of estimating the population mean of the variable under study using information on an auxiliary variable has received considerable attention in sampling from finite population. Ratio, product and linear regression estimators and their several generalizations which utilise information on a known population mean of an auxiliary variable, have been widely used in practice. A very large class of estimators of the population mean of the study variable using the known value of the population mean of the auxiliary variable was considered by Srivastava [7], [8]. Subsequently Srivastava and Jhajj [10] have defined a class of estimators which also utilizes known population variance of the auxiliary variable and it is shown that the lower bound for the asymptotic mean square error of this class of estimators is smaller than that of the class of estimators considered by Srivastava [7].

In this paper the class of estimators of Srivastava and Jhajj [10] is extended to a class which depends upon the known third moment of the auxiliary variable in addition to its mean and variance. Asymptotic expressions for the mean square error of an estimator of the class and also its minimum value are obtained. The expression by which the asymptotic minimum mean square error of estimator of this class is smaller than that which uses only the mean and variance of the auxiliary variable, is obtained. A similar class of estimators for the variance of the study variable is also considered. Numerical computation of efficiency of the proposed classes is made for six empirical populations taken from the literature.

2. *Notations*

Let Y_j and X_j, $j = 1, ..., N$ denote the values of the study variable y and the auxiliary variable x in the population. The corresponding lower case letters denote the value in the sample. We assume that a simple random sample of size n is drawn from the given finite population of size N. We write

$$\bar{y} = n^{-1} \sum_{j=1}^{n} y_j, \quad \bar{x} = n^{-1} \sum_{j=1}^{n} x_j,$$

$$\bar{Y} = N^{-1} \sum_{j=1}^{N} Y_j, \quad \bar{X} = N^{-1} \sum_{j=1}^{N} X_j,$$

$$s_x^2 = (n-1)^{-1} \sum_{j=1}^{n} (x_j - \bar{x})^2, \quad m_{x3} = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n} (x_j - \bar{x})^3,$$

$$\mu_{rs} = N^{-1} \sum_{j=1}^{N} (Y_j - \bar{Y})^r (X_j - \bar{X})^s,$$

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2} = \frac{\mu_{20}}{\bar{Y}^2}, \quad C_x^2 = \frac{S_x^2}{\bar{X}^2} = \frac{\mu_{02}}{\bar{X}^2},$$

$$\lambda_{rs} = \frac{\mu_{rs}}{S_y^r S_x^s}, \quad \rho = \lambda_{11}.$$

Let $u = \bar{x}/\bar{X}$, $v = s_x^2/S_x^2$ and $w = m_{x3}/\mu_{03}$, then we have :

$$E(\bar{y}/\bar{Y}) = E(s_y^2/S_y^2) = E(u) = E(v) = E(w) = 1,$$

$$E(\bar{y}/\bar{Y} - 1)^2 = n^{-1} C_y^2,$$

$$E(u-1)^2 = n^{-1} C_x^2,$$

$$E\{(\bar{y}/\bar{Y} - 1)(u-1)\} = n^{-1} \rho C_y C_x,$$

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

$$E(v-1)^2 = n^{-1} (\lambda_{04} - 1),$$

$$E\{(\bar{y}/\bar{Y} - 1)(v-1)\} = n^{-1} \lambda_{12} C_y$$

$$E\{(u-1)\ (v-1)\} = n^{-1}\lambda_{03}C_x,$$

$$E(w-1)^2 = n^{-1}(\lambda_{06} - 6\lambda_{04} - \lambda_{03}^2 + 9)/\lambda_{03}^2$$

$$E\{(\bar{y}/\bar{Y}-1)(w-1)\} = n^{-1}(\lambda_{13} - 3\rho)C_y/\lambda_{03}$$

$$E\{(u-1)\ (w-1)\} = n^{-1}(\lambda_{04} - 3)C_x/\lambda_{03}$$

$$E\{(v-1)\ (w-1)\} = n^{-1}(\lambda_{05} - 4\lambda_{03})/\lambda_{03}$$

$$E(s_y^2/S_y^2 - 1)^2 = n^{-1}(\lambda_{40} - 1)$$

$$E\{(s_y^2/S_y^2 - 1)(u-1)\} = n^{-1}\lambda_{21}C_x$$

$$E\{(s_y^2/S_y^2 - 1)(v-1)\} = n^{-1}\lambda_{22} - 1)$$

$$E\{(s_y^2/S_y^2 - 1)(w-1)\} = n^{-1}(\lambda_{23} - 3\lambda_{21} - \lambda_{03})/\lambda_{03}).$$

3. *The Class of Estimators of* $\bar{Y}$

Whatever the sample chosen, let (u, v, w) assume values in a bounded closed convex subset R, of the three dimensional real space containing the point (1, 1, 1). Let H (u, v, w) be a function of u, v and w such that

$$H(1,1,1) = 1 \tag{3.1}$$

and it satisfies the following conditions :

1. The function H (u, v, w) is continuous and bounded in R.
2. The first and second partial derivatives of H (u , v, w) exist and are continuous and bounded in R.

We consider the class of estimators of the population mean, $\bar{Y}$, defined by

$$\tilde{y}_H = \bar{y}\,H(u,v,w) \tag{3.2}$$

To find the bias and mean squre error of $\tilde{y}_H$ we expand the function H (u, v, w) about the point (1, 1, 1) in the second order Taylor's series. We have

$$\tilde{y}_H = \bar{y}\,H(u,v,w)$$

$$= \bar{y}\,[H(1,1,1) + (u-1)H_1(1,1,1) + (v-1)H_2(1,1,1)$$

$$+ (w-1)H_3(1,1,1) + \frac{1}{2}\{(u-1)^2 H_{11}(u^*, v^*, w^*)$$

$$+ (v-1)^2 H_{22}(u^*, v^*, w^*) + (w-1)^2 H_{33}(u^*, v^*, w^*)$$

$$+ 2(u-1)(v-1) H_{12}(u^*, v^*, w^*) + 2(u-1)(w-1) H_{13}(u^*, v^*, w^*)$$

$$+ 2(v-1)(w-1) H_{23}(u^*, v^*, w^*) \}] \quad (3.3)$$

where $u^* = 1 + \theta(u-1)$, $v^* = 1 + \theta(v-1)$ and $w^* 1 + \theta(w-1)$ $0 < \theta < 1$; $H_1(1,1,1)$, $H_2(1,1,1)$ and $H_3(1,1,1)$ denote the first order partial derivatives of $H(u, v, w)$ and $H_{11}(1,1,1)$, $H_{22}(1,1,1)$, $H_{33}(1,1,1)$, $H_{12}(1,1,1)$, $H_{13}(1,1,1)$, and $H_{23}(1,1,1)$, denote its second order partial derivatives with respect to u, v and w at the point (1,1,1).

Taking expectation and noting that the second order partial derivatives of H are bounded, and using (3.1), we have

$$E(\tilde{y}_H) = \overline{Y} + 0(n^{-1})$$

and so the bias of $\tilde{y}_H$ is of the order of n^{-1}. Using the results of section 2, the mean square error of $\tilde{y}_H$, upto terms of order n^{-1} is

$$M(\tilde{y}_H) = E(\tilde{y}_H - \overline{Y})^2$$

$$= E[(\bar{y} - \overline{Y})^2 + \bar{y}^2 \{(u-1)^2 H_1^2(1,1,1) + (v-1)^2 H_2^2(1,1,1)$$

$$+ (w-1)^2 H_3^2(1,1,1) + 2(\bar{y} - \overline{Y})(u-1) H_1(1,1,1) \bar{y}^{-1}$$

$$+ y^{-1}(\bar{y} - \overline{Y})(v-1) H_2(1,1,1) + 2(\bar{y} - \overline{Y})(w-1) H_3(1,1,1) \bar{y}^{-1}$$

$$+ 2(u-1)(v-1) H_1(1,1,1) H_2(1,1,1)$$

$$+ 2(u-1)(w-1) H_1(1,1,1) H_3(1,1,1)$$

$$+ 2(v-1)(w-1) H_2(1,1,1) H_3(1,1,1) \}]$$

$$= n^{-1} \overline{Y}^2 \{C_y^2 + C_x^2 H_1^2(1,1,1) + (\lambda_{04} - 1) H_2^2(1,1,1)$$

$$+ \frac{\lambda_{06} - 6\lambda_{04} - \lambda_{03}^2 + 9}{\lambda_{03}^2} H_3^2(1,1,1) + 2\rho C_y C_x H_1(1,1,1)$$

$$+ 2\lambda_{12} C_y H_2(1,1,1) + \frac{2(\lambda_{13} - 3\rho)}{\lambda_{03}} C_y H_3(1,1,1)$$

$$+ 2\lambda_{03} C_x H_1(1,1,1) H_2(1,1,1)$$

$$+\frac{2(\lambda_{04}-3)}{\lambda_{03}} C_x H_1(1,1,1) H_3(1,1,1)$$

$$\left. +\frac{2(\lambda_{05}-4\lambda_{03})}{\lambda_{03}} H_2(1,1,1) H_3(1,1,1)\right\} \quad (3.4)$$

The mean squre error of $\tilde{y}_H$ at (3.4) is minimized for

$$\begin{pmatrix} H_1(1,1,1) \\ H_2(1,1,1) \\ H_3(1,1,1) \end{pmatrix} = -C_y \underline{A}^{-1} \begin{pmatrix} \rho C_X \\ \lambda_{12} \\ \dfrac{\lambda_{13}-3\rho}{\lambda_{03}} \end{pmatrix} \quad (3.5)$$

where

$$\underline{A} = \begin{pmatrix} C_x^2 & \lambda_{03} C_x & \dfrac{(\lambda_{04}-3)C_x}{\lambda_{03}} \\ \lambda_{03} C_X & \lambda_{04}-1 & \dfrac{(\lambda_{06}-4\lambda_{03})}{\lambda_{03}} \\ \dfrac{(\lambda_{04}-3)C_x}{\lambda_{03}} & \dfrac{(\lambda_{05}-4\lambda_{03})}{\lambda_{03}} & \dfrac{(\lambda_{06}-6\lambda_{04}-\lambda_{03}^2+9)}{\lambda_{03}^2} \end{pmatrix} \quad (3.6)$$

Explicit expressions of $H_1(1,1,1)$, $H_2(1,1,1)$ and $H_3(1,1,1)$ in (3.5) are given below.

$$H_1(1,1,1) = \frac{C_y}{C_x} [\{\lambda_{12}\lambda_{03} - \rho(\lambda_{04}-1)\}(\lambda_{06}-6\lambda_{04}-\lambda_{03}^2+9)$$

$$-\{(\lambda_{04}-1)(\lambda_{04}-3) - \lambda_{03}(\lambda_{05}-4\lambda_{03})\}(\lambda_{13}-3\rho)$$

$$-\{(\lambda_{12}(\lambda_{04}-3) - \rho(\lambda_{05}-4\lambda_{03}))\}(\lambda_{05}-4\lambda_{03})]/\lambda_{03}^2 K$$

$$H_2(1,1,1) = C_y [(\lambda_{03}-\lambda_{12})(\lambda_{06}-6\lambda_{04}-\lambda_{03}^2+9)$$

$$-\{\lambda_{03}(\lambda_{04}-3) - (\lambda_{05}-4\lambda_{03})\}(\lambda_{13}-3\rho)$$

$$+\{\lambda_{12}(\lambda_{04}-3) - \rho(\lambda_{05}-4\lambda_{03})\}(\lambda_{04}-3)]/\lambda_{03}^2 K$$

$$H_3(1,1,1) = C_y [\{\rho(\lambda_{04}-1) - \lambda_{12}\lambda_{03}\}(\lambda_{04}-3)$$

$$-(\lambda_{04}-\lambda_{03}^2-1)(\lambda_{13}-3\rho) - (\rho\lambda_{03}-\lambda_{12})(\lambda_{05}-4\lambda_{03})]/\lambda_{03} K$$

where

$$K = (\lambda_{06} - \lambda_{04}^2 - \lambda_{03}^2)(\lambda_{04} - \lambda_{03}^2 - 1) - (\lambda_{05} - \lambda_{04}\lambda_{03} - \lambda_{03})^2$$

The minimum mean square error of $\tilde{y}_H$ is given by

$$\min M(\tilde{y}_H) = \frac{S_y^2}{n}\left[(1-\rho^2) - \frac{(\rho\lambda_{03} - \lambda_{12})^2}{\lambda_{04} - \lambda_{03}^2 - 1} - \frac{\{(\rho\lambda_{04} - \lambda_{13})(\lambda_{04} - \lambda_{03}^2 - 1) + (\lambda_{12} - \rho\lambda_{03})(\lambda_{05} - \lambda_{04}\lambda_{03} - \lambda_{03})\}^2}{(\lambda_{04} - \lambda_{03}^2 - 1)\{(\lambda_{05} - \lambda_{04}^2 - \lambda_{03}^2)(\lambda_{04} - \lambda_{03}^2 - 1) - (\lambda_{05} - \lambda_{04}\lambda_{03} - \lambda_{03})^2\}}\right] \quad (3.7)$$

The third term on the right hand side of (2.7) gives the amount of reduction in the minimum asymptotic mean square error of the class when μ_{03} is also used in addition to $\overline{X}$ and S_x^2. The gain in efficiency by the use of the knowledge of μ_{03} is illustrated for six empirical populations in section 5.

The class (3.2) of estimators is very large. The following function, for example give some simple estimators of the class :

$$H(u, v, w) = u^\alpha v^\beta w^\gamma,$$

$$H(u, v, w) = a_1 u^\alpha + a_2 v^\beta + a_3 w^\gamma,\ a_1 + a_2 + a_3 = 1,$$

$$H(u, v, w) = 1 + \alpha(u-1) + \beta(v-1) + \gamma(w-1),$$

$$H(u, v, w) = \{1 + \alpha(u-1) + \beta(v-1)\}\ \{1 - \gamma(w-1)\}^{-1}$$

$$H(u, v, w) = \{1 - \alpha(u-1) - \beta(v-1) - \gamma(w-1)\}^{-1}.$$

The optimum values of the unknown parameters α, β and γ in these estimators are determined from the condition (3.5), and with these optimum values the mean square error up to terms or order n^{-1}, is given by (3.7). In practice, the optimum values of these parameters will involve unknown population parameters such as ρ, λ_{12} and λ_{13}. However, if consistent estimators of these population parametes are used in the optimum values obtained from (3.5) there will be change in the mean square error in terms which are $0(n^{-2})$. And so the mean square error up to terms of order n^{-1}will remain unaltered.

Following Srivastava (1980), it is easily shown that even for the wider class of estimators

$$\tilde{y}_G = G(\bar{y}, u, v, w)$$

of $\overline{Y}$, where the function G satisfies $G(\overline{Y}, 1, 1, 1) = \overline{Y}$, the minimum asymptotic mean square error is same as (3.7) and is not less.

4. *The Class of Estimators of* S_y^2

Srivastava and Jhajj (1980) defined a class of estimators for the population variance S_y^2 of the study variable y utilising the known value of $\overline{X}$ and S_x^2. Following the approach of the preceeding section, we extend this class of estimators of S_y^2 to

$$\hat{t}_H = s_y^2 H(u, v, w), \tag{4.1}$$

where the function H (u, v, w) again satisfies the conditions of section 3.

Following the method of section 3, the bias of $\hat{t}_H$ is seen to be of order n^{-1} and its mean square error up to terms of order n^{-1} is minimixed for

$$\begin{pmatrix} H_1(1,1,1) \\ H_2(1,1,1) \\ H(1,1,1) \end{pmatrix} = -\underline{A}^{-1} \begin{pmatrix} \lambda_{21} C_x \\ \lambda_{22} - 1 \\ \dfrac{\lambda_{23} - 3\lambda_{21} - \lambda_{03}}{\lambda_{03}} \end{pmatrix}$$

where $\underline{A}$ is the matrix defined by (3.6). The minimum value of the mean square error of $\hat{t}_H$ is given by

$$\min M(\hat{t}_H) = \frac{S_y^4}{n}\left[(\lambda_{40}-1) - \frac{(\lambda_{22}-1)^2}{\lambda_{04}-1} - \frac{\{\lambda_{21}(\lambda_{04}-1) - \lambda_{03}(\lambda_{22}-1)\}^2}{(\lambda_{04}-1)(\lambda_{04}-\lambda_{03}^2-1)} - \frac{\{(\lambda_{23}-\lambda_{21}\lambda_{04}-\lambda_{03})(\lambda_{04}-\lambda_{03}^2-1) + (\lambda_{21}\lambda_{03}-\lambda_{22}+1)(\lambda_{05}-\lambda_{04}\lambda_{03}-\lambda_{03})\}^2}{(\lambda_{04}-\lambda_{03}^2-1)\{(\lambda_{06}-\lambda_{04}^2-\lambda_{03}^2)(\lambda_{04}-\lambda_{03}^2-1) - (\lambda_{05}-\lambda_{04}\lambda_{03}-\lambda_{03})^2\}}\right] \tag{4.2}$$

In (4.2) the first term on the right hand side gives the asymptotic mean square error when no auxiliary information is used, the second term gives the reduction in mean square error when S_x^2 is used, the third term gives the reduction when $\overline{X}$ is used along with S_x^2 and the fourth term gives the reduction when μ_{03} is used along with $\overline{X}$ and S_x^2.

5. *Numerical Illustration*

To have an idea of the increase in efficiency of optimum estimators of the classes (3.2) and (4.1) or $\overline{Y}$ and S_y^2 respectively, by using the third moment of the auxiliary variable in addition to its mean and variance, we have made computatioins for six natural populations taken from literature. The source of the population and the nature of the variable x and y are given in Table 5.1. In Table 5.2 we have listed the efficiencies of optimum estimators of the proposed class using μ_{03} along with $\overline{X}$ and S_x^2

Table 5.1. Description of populations

S.No	*Source*	*y*	*x*
1.	Cochran [1], p, 325	No. of persons per block	No. of rooms per block
2.	Horvitz and Thompson [2]	No. of households	Eye estimate of y
3.	Cochran [1], p. 203	Actual weight of peaches on each tree	Eye estimate of weight of peaches on each tree
4.	Sukhatme and Sukhatme [11] p.185, vil.1-10.	Wheat aereage in 1937.	Wheat acreage in 1936
5.	Murthy [4] p.399, vil.1-10	Area under wheat in 1964	Area under wheat in 1963.
6.	Cochran [1] p. 152, cities 1-16	No. of inhabitants in 1930	No. of inhabitants in 1920.

Table 5.2 has been computed from the expressions (3.7) and (4.2). The efficiency has been taken to be the ratio of the reciprocals of the mean square errors. The mean square error of an estimator of $\overline{Y}$ and $\overline{X}$ alone has been computed by taking the first term only on the r.h.s of (3.7). For the mean squre using $\overline{X}$ and S_x^2, the first two terms on the r.h.s. of (3.7) have been taken and for the mean square error using $\overline{X}$, S_x^2 and μ_{03}, all the three terms on the r.h.s. of (3.7) have been taken. The computation of mean square error for estimators of S_y^2 have similarly been done from the expression (4.2).

It is observed from Table 5.2 that the use of the knowledge of μ_{03}, the third moment of the auxiliary variable, results in substantial increase in the efficiency for the estimation of $\overline{Y}$ in case of two out of six populations considered. In the remaining four populations the increase in efficiency is only marginal. For the estimation of S_y^2, however, the use of the knowledge of μ_{03} results in substantial increase in the efficiency in the case of all the six populations.

Table. 5.2. Percentage efficiency of optimum estimators of $\overline{Y}$ and S_y^2 for different populations of Table 5.1.

S.No. of the population.	*Optimum estimators of $\overline{Y}$*		
	Using $\overline{X}$ alone	Using $\overline{X}$ and S_x^2	Using $\overline{X}$, S_x^2 and μ_{03}
1.	100	111.21	112.61
2.	100	103.11	116.87
3.	100	175.74	199.24
4.	100	115.95	280.47
5.	100	116.26	266.25
6.	100	305.13	324.61
	Optimum estimators of S_y^2		
	Using S_x^2 alone	Using $\overline{X}$ and S_x^2	Using $\overline{X}$, S_x^2 and μ_{03}
1.	100	106.00	148.00
2.	100	103.34	131.96
3.	100	136.76	630.87
4.	100	100.71	235.14
5.	100	100.86	241.12
6.	100	114.88	157.60

REFERENCES

[1] Cochran, W.C., 1977. *Sampling Techniques*. New York : John Wiley and Sons.

[2] Horvitz, D.G. and Thompson, D.J., 1952."A Generalization of Sampling without Replacement from a Finite Universe". *Journal of the American Statistical Association*, **47**, 663-685.

[3] Jhajj, H.S., 1982. *Contributions to the use of auxiliary information* in survey sampling". unpublished Ph.D. thesis, Punjabi University, Patiala.

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

[5] Nath, S.N., 1969. More Results on Product Moments from a Finite Universe. *Journal of the American Statistical Association*, **64** : 864-869.

[6] Raghunandan, K. and Srinivasan, R., 1973. "Some Product Moments useful in Sampling Theory." *Journal of the American Statistical Association*, **68** : 409-413.

[7] Srivastava, S.K., 1971. A Generalized Estimator for the Mean of a Finite Population using multi-Auxiliary Information. *Journal of the American Statistical Association*, **66**. 404-407.

[8] Srivastava, S.K., 1980. A Class of Estimators Using Auxiliary Information in Samplign Surveys. *Candian Journal of Statistics*, **8**. 253-254.

[9] Srivastava, S.K. and Jhajj, H.S., 1980. A class of Estimators using Auxiliary Information for Estimating Finite Population Variance. *Sankhya, Ser.C.*, **42**. 87-96.

[10] Srivastava, S.K. and Jhajj, H.S., 1981. A class of Estimators of the Population Mean in Survey Sampling using Auxiliary Information. *Biometrike*, **68**. 341-343.

[11] Sukhatme, P.V. and Sukhatme, B.V., 1970. *Sampling Theory of Surveys with Applications*. New Delhi : Asia Publishing House.