Jour. Ind. Soc. Ag. Statistics
*Vol. XXXVIII, No. 1 (1986), pp.*100-104

ON A CLASS OF ESTIMATORS OF THE POPULATION MEAN IN SURVEY SAMPLING USING AUXILIARY INFORMATION

H. P. SINGH* and L. N. UPADHYAYA
Indian School of Mines, Dhanbad

(Received : July, 1984)

SUMMARY

This paper proposes a class of estimators of the population mean wider as well as more efficient than those considered by Srivastava and Jhajj [3] using auxiliary information on a character x. Asymptotic expressions for bias and mean squared error are obtained.

Keywords : Finite population; Linear regression estimator; asymptotic Variance.

1. Introduction and Notation

Let $U = (1, 2, \ldots, N)$ be a finite population of N units and y be a real valued function defined on U taking the value y_i for the unit i of U $(1 \leqslant i \leqslant N)$. Let x be an auxiliary variate correlated with y, taking the value x_i on unit $i(1 \leqslant i \leqslant N)$, information on which is available in advance.

Further, we assume that a simple random sample of size n is drawn from U. For simplicity we assume that N is large as compared to n so that finite population correction terms are ignored. We write

$$\overline{Y} = N^{-1} \sum_{i=1}^{N} y_i,\ \overline{X} = N^{-1} \sum_{i=1}^{N} x_i,\ \mu_{rs} = N^{-1} \sum_{i=1}^{N} (y_i - \overline{Y})^r (x_i - \overline{X})^s$$

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

*Work done as CSIR Fellow at Indian School of Mines, Dhanbad.
Present Address : Assistant Prof. (statistics), JNK Vishwavidyalaya, Jabalpur.

$$\lambda = \frac{\mu_{12}}{Y S_x^2}, \gamma_1 = \frac{\mu_{03}}{S_x^3}, \beta_2 = \frac{\mu_{04}}{S_x^4}, \bar{y} = \sum_{i=1}^{n} y_i/n,$$

$$\bar{x} = \sum_{i=1}^{n} x_i/n, s_x^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \beta_1 = \gamma_1^2 .$$

Let $W = \bar{y}/\bar{Y}$, $u = \bar{x}/\bar{X}$ and $v = s_x^2/S_x^2$. Then we have

$$E(W) = E(u) = E(v) = 1, E\{(W-1)^2\} = n^{-1} C_y^2,$$
$$E\{(u-1)^2\} = n^{-1} C_x^2, E\{(W-1)(u-1)\} = n^{-1} \rho C_y C_x$$

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

$$E\{(v-1)^2\} = n^{-1} (\beta_2 - 1), E\{(W-1)(v-1)\} = n^{-1}\lambda,$$
$$E\{(u-1)(v-1)\} = n^{-1} \gamma_1 C_x.$$

Following Srivastava [2], Srivastava and Jhajj [1] considered a class of estimators for $\bar{Y}$, defined by

$$\bar{y}_t = \bar{y}\, t(u, v), \tag{1.1}$$

where $t(u, v)$ is a function of u and v which satisfies certain conditions.

Further, they improved the estimator $\bar{y}_t$ in (1.1) in their same paper and defined another class of estimators wider than $\bar{y}_t$ as

$$\bar{y}_g = g(\bar{y}, u, v) \tag{1.2}$$

where $g(\bar{y}, u, v)$ is a function of $\bar{y}$, u and v such that

(i) $(\bar{y}, u, v)$ assume values in a closed convex, subset, P, of three dimensional real space containing the point $(\bar{Y}, 1, 1)$,

(ii) The function $g(\bar{y}, u, v)$ is a continuous and bounded in P,

(iii) $g(Y, 1, 1) = \bar{Y}$, (1.3)

(iv) The first and second order partial derivatives of $g(\bar{y}, u, v)$ exist and are continuous and bounded in P.

The minimum mean squared error of both the classes of estimators $\bar{y}_t$ and $\bar{y}_g$ defined in (1.1) and (1.2) respectively, are same as given by

$$\min M(\bar{y}_z) = n^{-1} \bar{Y}^2 \left[C_y^2 (1 - \rho^2) - \frac{(\rho C_y \gamma_1 - \lambda)^2}{(\beta_2 - \beta_1 - 1)} \right] \tag{1.4}$$

$$z = t, g.$$

No doubt, these two estimators $\bar{y}_z$, $z = t, g$ considered by Srivastava and Jhajj (3) are too vast, but these classes of estimators fail to include a simple class of estimators suggested by Searl's [1] defined by

$$\bar{y}_s = W\bar{y}, \tag{1 5}$$

where W is a suitably chosen constant.

The minimum MSE of $\bar{y}_s$ is given by

$$\min M(\bar{y}_s) = \bar{Y}^2 C_y^2/(n + C_y^2). \tag{1.6}$$

The estimators suggested by Srivastava and Jhajj [3] also suffer from a drawback that 'in case of bivariate normal population, there is no contribution of $v(= s_x^2/S_x^2)$. In such a population the minimum MSE of $\bar{y}_z$, $z = t, g$ reduces to that of the usual linear regression estimator $\bar{y}_{1r} = \bar{y} + b\,(\bar{X} - \bar{x})$, b being the sample regression coefficient of y on x which is given by

$$M(\bar{y}_{1r}) = n^{-1}\,\bar{Y}^2\,C_y^2\,(1 - \rho^2) \tag{1.7}$$

In this paper we suggest a class of estimators of $\bar{Y}$, using information on an auxiliary character x, wider as well as more efficient than those considered by Srivastava and Jhajj [3] and include the estimator $\bar{y}_s$ suggested by Searls [1] and other estimators possibly constructed on the lines of Searls [1]. The present estimator is also more efficient than Srivastava and Jhajj [3] estimators, Searls [1] estimator, usual linear regression estimator $\bar{y}_{1r}$ and unbiased estimator $\bar{y}$, even in case of a bivariate normal population.

2. The Class of Estimators

We propose a class of estimators for the population mean $\bar{Y}$ as

$$\bar{y}_h = h\,(\bar{y}, u, v), \tag{2.1}$$

where $h(\bar{y}, u, v)$ is a function of $\bar{y}$, u and v such that

(i) $(\bar{y}, u, v)$ assume the values in a bounded closed convex subset, P, of three dimensional real space containing the point $(\bar{Y}, 1, 1)$;

(ii) $h(\bar{y}, u, v)$ is a continuous and bounded in P;

(iii) $h(\bar{Y}, 1, 1) = \bar{Y}\,h_1(\bar{Y}, 1, 1), \ldots,$ (2.2)

where $h_1(\bar{Y}, 1, 1)$ denotes the first order partial derivative with respect to $\bar{y}$ at $(\bar{y}, u, v) = (\bar{Y}, 1, 1)$;

(iv) The first and Second order partial derivatives of $h(\bar{y}, u, v)$ exist and are continuous and bounded in P.

It is to be noted here that $h_1(\overline{Y}, 1, 1)$ is a constant (which may, in particular, be unity). It is also interesting to remark that the class of estimators $\bar{y}_h = h(y, u, v)$ reported here reduces to the class of estimators $\bar{y}_g = g(\bar{y}, u, v)$ forwarded by Srivastava and Jhajj [3] for $h_1(\overline{Y}, 1, 1) = 1$.

Expanding $h(\bar{y}, u, v)$ about the point $(\overline{Y}, 1, 1)$ in a second order Taylor's series, we have that $E(\bar{y}_h) = \overline{Y} h_1 (\overline{Y}, 1, 1) + 0(n^{-1})$, and also the bias of $\bar{y}_h$ is of the order of n^{-1}. The mean squared error of $\bar{y}_h$ upto terms of order n^{-1} is

$$\begin{aligned} M(\bar{y}_h) = \; & [\overline{Y}^2 \{1 - h_1 (\overline{Y}, 1, 1)\}^2 + n^{-1} \{\overline{Y}^2 C_y^2 h_1^2 (\overline{Y}, 1, 1) \\ & + C_x^2 h_2^2 (\overline{Y}, 1, 1) + (\beta_2 - 1) h_3^2 (\overline{Y}, 1, 1) \\ & + 2\rho C_y C_x h_1(\overline{Y}, 1, 1) h_2(\overline{Y}, 1, 1) + 2\lambda \overline{Y} h_1(\overline{Y}, 1, 1) \\ & h_3(\overline{Y}, 1, 1) + 2\gamma_1 C_x h_2(\overline{Y}, 1, 1) h_3(\overline{Y}, 1, 1)\}] \end{aligned} \tag{2.3}$$

where $h_i(\bar{y}, u, v)$; $i = 1, 2, 3$ denote the first order partial derivatives of $h(\bar{y}, u, v)$. The mean squared error of $\bar{y}_h$ at (2.3) is minimized for

$$\begin{aligned} h_1(\overline{Y}, 1, 1) &= \frac{\Delta_1}{\Delta}, \\ h_2(\overline{Y}, 1, 1) &= -(R\overline{X}/C_x) \cdot \frac{\Delta_2}{\Delta}, \\ h_3(\overline{Y}, 1, 1) &= R\overline{X} \cdot \frac{\Delta_3}{\Delta}, \end{aligned} \tag{2.4}$$

where $\Delta_1 = (\beta_2 - \beta_1 - 1)$, $\Delta_2 = [C_y(\beta_2 - 1) - \lambda\gamma_1]$, $\Delta_3 = (\rho C_y \gamma_1 - \lambda)$, $R = \overline{Y} / \overline{X}$ and $\Delta = [(\beta_2 - \beta_1 - 1) \{1 + n^{-1} C_y^2 (1 - \rho^2)\} - n^{-1} (\rho C_y \gamma_1 - \lambda)^2]$. Hence the resulting (minimum) mean squared error of $\bar{y}_h$ is given by

$$\min M(\bar{y}_h) = n^{-1}\overline{Y}^2 [(\beta_2 - \beta_1 - 1) C_y^2 (1 - \rho^2) - (\rho C_y \gamma_1 - \lambda)^2]/\Delta. \tag{2.5}$$

From (1.4) and (2.5) we have

$$\min M(\bar{y}_g) - \min M(\bar{y}_h) = n^{-1}\overline{Y}^2 \frac{[(\beta_2 - \beta_1 - 1) C_y^2(1 - \rho^2) - (\rho C_y \gamma_1 - \lambda)^2]^2}{(\beta_2 - \beta_1 - 1)} \tag{2.6}$$

which is always positive. Hence the minimum MSE of proposed class of estimators $\bar{y}_h$ is always less than the minimum MSE's of the estimators reported by Srivastava and Jhajj [3], Searls [1], the linear regression estimator $\bar{y}_{1r}$ and the usual unbiased estimator $\bar{y}$.

Srivastava and Jhajj [3] remarked that the minimum mean squared error of their class of estimators would be less than that of the following class of estimators

$$\bar{y}_f = \bar{y} f(u); \tag{2.7}$$

suggested by Srivastava [2], where $f(u)$ is a function of u which satisfies certain regularity conditions [see Srivastava [2] pp. 405], if and only if $\rho C_y \gamma_1 \neq \lambda$. But the minimum mean squared error of our proposed class of estimators is less than that of $\bar{y}_f$ even when $\rho C_y \gamma_1 = \lambda$. Further, in case of a bivariate normal population the minimum MSE of $\bar{y}_h$ reduces to

$$\min M(\bar{y}_h) = n^{-1} \bar{y}^2 C_y^2 (1 - \rho^2)/[1 + n^{-1} C_y^2 (1 - \rho^2)], \tag{2.8}$$

which is less than the asymptotic variance of the estimators $\bar{y}_{1r}$, $\bar{y}_t$, $\bar{y}_a$, $\bar{y}_s$ and $\bar{y}$.

ACKNOWLEDGEMENTS

The authors are thankful to the referee for valuable comments for the improvement of the paper. One of the authors Dr. H. P. Singh is thankful to CSIR also for the financial support.

REFERENCES

[1] Searls, D. T. (1964) : The utilization of a known coefficient of variation in the estimation procedure, *Jour. Amer. Statist. Assoc.*, **59** : 1225-1226.

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

[3] Srivastava, S. K. and Jhajj, H. S. (1981) : A class of estimator of the population mean in surveys sampling using auxiliary information, *Biometrika*, **68**(1): 341-43.