A CLASS OF PPS ESTIMATORS OF POPULATION MEAN USING AUXILIARY INFORMATION

BY

HARBANS SINGH JHAJJ AND SURENDRA K. SRIVASTAVA

Punjabi University, Patiala

(Received : September, 1980)

SUMMARY

A class of estimators for the mean of a finite population has been considered using information on two auxiliary variables, one of which is used at the sample selection stage and other for improving the estimator at the estimation stage. This generalizes the estimator proposed by Agarwal and Kumar (1980). The asymptotic bias and mean square error of estimators in the class have been obtained. Also the optimum estimators of the class have been obtained.

Srivastava and Jhajj [3] have defined estimators which utilize the value of the population variance of the auxiliary variable and have shown that the resulting estimators have smaller asymptotic mean square error than that of the linear regression estimator. Recently Agarwal and Kumar [1] used two auxiliary variables : one at the stage of selection of the sample and the other at the estimation stage and then taking the best linear combination of the probability proportional to size (PPS) estimator and the ratio estimator so obtained, to estimate the population mean of the study variable. The minimum asymptotic mean square error of the proposed estimator is shown to have the same form as the variance of the regression estimator under PPS with replacement sampling.

Following Srivastava and Jhajj [3] a class of estimators of the population mean of the study variable has been defined when the sampling is done by the method of probability proportional to a suitable size variable which is different from the auxiliary variable used at estimated stage. Asymptotic expressions for the bias and mean square error of the proposed class of estimators are obtained.

$$\sigma_u^2=\sum_{i=1}^{n} P_i(u_i-\bar{Y})^2, \quad \sigma_v^2=\sum_{i=1}^{N} P_i(v_i-\bar{X})^2$$

$$c_u^2=\frac{\sigma_u^2}{\bar{Y}^2}; \quad c_v^2=\frac{\sigma_v^2}{\bar{X}^2}$$

$$\lambda rs=\sum_{i=1}^{N} P_i (u_i-\bar{Y})^r (v_i-\bar{X})^s.$$

$$\rho_{uv}=\frac{\lambda_{11}}{\sigma u \sigma v}, \quad \psi=\frac{\lambda_{12}}{\sigma u \; v_u^2}$$

$$\theta_2 \frac{\lambda_{04}}{\sigma_u^4}, \quad \alpha_1=\frac{\lambda_{03}}{\sigma_v^3}, \quad \theta_1=\alpha_1^2.$$

Defining

$$w=\frac{\bar{v}_n}{\bar{X}}, \quad z=\frac{s_v^2}{\sigma_v^2}$$

and

$$\varepsilon=\frac{\bar{u}_n}{\bar{Y}}-1, \quad \delta=w-1, \quad \eta=z-1,$$

we have

$$E(\varepsilon)=E(\delta)=E(\eta)=0$$

$$E(\varepsilon^2)=\frac{1}{n} C_u^2, \quad E(\delta^2)=\frac{1}{n} C_v^2$$

$$E(\eta^2)=\frac{1}{n}\left\{\theta_2-\left(\frac{n-3}{n-1}\right)\right\}, \quad E(\varepsilon\,\delta)=\frac{1}{n} \rho_{uv} C_u C_v$$

$$E(\varepsilon\eta)=\frac{1}{n} \psi\, Cu, \quad E(\delta\,\eta)=\frac{1}{n} \alpha_1 C_v$$

Notations

Suppose that information on two auxiliary variables highly correlated with the study variable y is available. Let a sample of size n be drawn with PPS sampling and with replacement. Let P_i denote the probability of selection (based on one of the two auxiliary variables) of its unit, $i=i, \ldots, N$. Let y_i and x_i denote the value of the variable under study y and the auxiliary variable x for the ith unit of the population, and $\bar{Y}$ and $\bar{X}$ denote their population means respectively. We write.

$$u_i = \frac{y_i}{NP_i}, \quad v = \frac{x_i}{NP_i},$$

$$\bar{u}_n = \frac{1}{n} \sum_{i=1}^{n} u_i \; ; \; \bar{v}_n = \frac{1}{n} \sum_{i=1}^{n} v_i$$

$$s_v^2 = \frac{1}{n\text{-}1} \sum_{i=1}^{n} (v_i\text{-}\bar{v}_n)^2$$

I THE CLASS OF ESTIMATORS AND MEAN SQUARE ERROR

Consider the PPS sampling scheme with replacement based on P_i. Then the proposed class of estimators of $\bar{Y}$ is

$$T = \bar{u}_n \, t\,(w,z) \qquad \ldots(3.1)$$

where $t\,(w,z)$ is a function of w and z such that

$$t\,(1,1) = 1. \qquad \ldots(3.2)$$

We assume that the function $t\,(w,z)$ is continuous and has continuous first and second partial derivatives which are bounded in a closed convex subset, D, of the two dimensional real space containing the point (1,1).

Expanding the function t (w,z) about the point (1,1) in a second-order Taylor's series, we have

$$T = \bar{u}_n \, [t\,(1,1) + (w\text{-}1)\, t_1\,(1,1) + (z\text{-}1)\, t_2\,(1,1)$$
$$+ \frac{1}{2}\{(w\text{-}1)^2\, t_{11}\,(w^*, z^*) + 2\,(w\text{-}1)\,(z\text{-}1)\, t_{12}\,(w^*, z^*)$$
$$+ (z\text{-}1)^2\, t_{22}\,(w^*, z^*)\,\}\,] \qquad \ldots(3.3)$$

where $w^* = 1 + \theta\,(w\text{-}1)$, $z^* = 1 + \theta(z\text{-}1)$, $0 < \theta < 1$ and. $t_1\,(w, z)$, $t_2\,(w, z)$ denote the first partial derivatives of the function $t\,(w, z)$ at the point (w, z) and $t_{11}\,(w^*, z^*)$, $t_{12}\,(w^*, z^*)$ and $t_{22}\,(w^*, z^*)$ denote its second partial derivatives at the point (w^*, z^*).

Substituting for $\bar{u}_n$, w and z in terms of ε, δ and η in (3.3) and using (3.2), we have

$$T = \bar{Y}[1+\varepsilon+\delta t_1(1,1)+\eta t_2(1,1)+\delta\varepsilon t_1(1,1+\eta\varepsilon t_2(1,1)$$
$$+\tfrac{1}{2}\{\delta^2 t_{11}(w^*,z^*)+2\delta\eta t_{12}(w^*,z^*)+\eta^2 t_{22}(w^*,z^*)\}] \qquad \ldots(3.4)$$

Taking expectation in (3.4), it is easily found that

$$E(T) = \bar{Y} + 0(n^{-1}).$$

The mean square error of T up to terms of order n^{-1} is given by

$$M(T)=E(T-\bar{Y})^2$$

$$=\bar{Y}^2\{E\varepsilon^2+\delta^2 t_1^2(1,1)+\eta^2 t_2^2(1,1)+2\varepsilon\delta t_1(1,1)+2\varepsilon\eta t_2(1,1)+2\delta\eta t_1(1,1)\, t_2(1,1)\}$$

$$=\frac{\bar{Y}^2}{n}\{C_u^2+C_v^2 t_1^2(1,1)+(\theta_2-1)t_2^2(1,1)+2\rho_{uv}C_uC_v t_1(1,1)+2\Psi C_u t_2(1,1)+2\alpha_1 C_v t_1(1,1)\, t_2(1,1)\} \quad ...(3.5)$$

The optimum values of $t_1(1,1)$ and $t_2(1,1)$ for which the mean square error $M(T)$ at (3.5) is minimised is given by

$$t_1(1,1)=\frac{C_u}{C_v}\cdot\frac{\{\Psi\alpha_1-\rho_{uv}(\theta_2-1)\}}{(\theta_2-\theta_1-1)} \quad ...(3.6)$$

$$t_2(1,1)=\frac{C_u(\rho_{uv}\alpha_1-\Psi)}{\theta_2-\theta_1-1} \quad ...(3.7)$$

Substituting from (3.6) and (3.7) in (3.5), the minimum mean square error of T, up to terms of order n^{-1}, is given by.

$$\text{Min } M(T)=\frac{1}{n}\bar{Y}^2C_u^2\left\{1-\rho_{uv}^2-\frac{(\alpha_1\rho_{uv}-\Psi)^2}{\theta_2-\theta_1-1}\right\}. \quad ...(3.8)$$

Since $\theta_2-\theta_1-1>0$, it follows from (3.8) that the minimum mean square error of any estimator of the class (3.1), up to terms of order n^{-1}, is not greater than the minimum mean square error of the estimator proposed by Agarwal and Kumar (1980). In fact the second term on the right of (3.8) gives an idea of the amount of the decrease in the mean square error. The class (3.1) of estimators is very large. Any parametric function $t(w,z)$ satisfying (3.2) can generate an estimator of the class. If the parameters in $t(w,z)$ are so chosen that they satisfy (3.6) and (3.7), then the resulting estimator will have the asymptotic mean square error given by (3.8). A few examples of the function $t(w,z)$ are given in Srivastava and Jhajj [3]

2. THE BIAS

To obtain the bias of the estimator (3.1), it is further assumed that third order partial derivatives of the function $t(w,z)$ exist and are continuous and bounded in D. Then, expanding the function $t(w,z)$ in a third order Taylor's series about the point (1,1), the bias of T is detained.

$$\begin{aligned}\text{Bias}(T) &= E(T-\bar{Y})\\ &= \bar{Y}[t_1(1,1)\,E(\delta\varepsilon)+t_2(1,1)\,E(\varepsilon\eta)+\tfrac{1}{2}\{t_{11}(1,1)\,E(\delta^2)\\ &\quad +2t_{12}(1,1)\,E(\delta\eta)+t_{22}(1,1)\,E(\eta^2)\}]\\ &= \frac{\bar{Y}}{2n}\{2t_1(1,1)\,\rho_{uv}C_uC_v+2\Psi C_u t_2(1,1)+2a_1C_v t_{12}(1,1)\\ &\quad +C_v^2 t_{11}(1,1)+(\theta_2-1)t_{22}(1,1)\} \qquad (4.1)\end{aligned}$$

where $t_{11}(1,1)$, $t_{12}(1,1)$ and $t_{22}(1,1)$ denote the second partial derivatives of the function $t(w,z)$ at the point $(1,1)$. Thus we see that the bias of the estimator T also depends upon the second partial derivatives of the function $t(w,z)$ at the point $(1,1)$ and hence will be different for different optimum estimators of the class. From (4.1), one can easily obtain the asymptotic bias of any estimator of the class (3.1). It is interesting to note that if we take

$$t(w,z)=\frac{1}{1-\alpha(w-1)-\beta(z-1)}$$

in defining an estimator T of the class (3.1), the asymptotic bias of this estimator T with optimum values of the parameters α and β, is equal to zero.

References

[1] Agarwal S.K. and Kumar P (1980) : Combination of ratio and PPS estimator, *Jour. Indian Society of Agricultural Statistics* **32**, pp. 81-86.

[2] Murthy M.N. (1967) : *Sampling theory and Methods* Statistical publishing Society Calcutta. pp. 399.

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