Jour. Ind. Soc. Ag. Statistics
47(2), 1995 : 168–175

A Class of Unbiashed Dual to Ratio Estimator in Stratified Sampling

Housila P. Singh and Virendra P. Singh*
Vikram University, Ujjain, M.P.
(Received : May, 1993)

SUMMARY

This paper proposes a class of unbiased dual to ratio estimator for population mean and analyses its properties.

Key words : Dual to ratio estimator, Positively correlated auxiliary variables, Optimum estimator, Variance.

Introduction

Assume that the population of size N is divided into L strata and that sampling within each stratum is simple random sampling without replacement (SRSWOR). Let N_h denote the number of units in the h–th stratum and n_h the size of the sample to be selected therefrom, so that

$$\sum_{h=1}^{L} N_h = N \text{ and } \sum_{h=1}^{L} n_h = n.$$

Let $$\bar{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{j=1}^{N_h} y_{hj} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{Y}_h = \sum_{h=1}^{L} P_h \bar{Y}_h \text{, and}$$

$$\bar{X} = \frac{1}{N}\sum_{h=1}^{L}\sum_{j=1}^{N_h} x_{hj} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{X}_h = \sum_{h=1}^{L} P_h \bar{X}_h$$

be the population means of positively correlated characters y (study) and x (auxiliary) respectively, where $P_h = N_h/N$, $\bar{Y}_h = \sum_{j=1}^{N_h} y_{hj}/N_h$ and $\bar{X}_h = \sum_{j=1}^{N_h} \frac{x_{hj}}{N_h}$. To estimate $\bar{Y}$, the usual separate ratio estimator is defind by

$$\hat{\bar{Y}}_{RS} = \sum_{h=1}^{L} P_h \bar{y}_{Rh} \tag{1.1}$$

* Gujarat Agricultural University, Anand Campus, Anand–388110.

where, $$\bar{y}_{RH} = \bar{y}_h \, (\bar{X}_h/\bar{x}_h) \tag{1.2}$$

is the ratio estimator of $\bar{Y}_h$, the population mean for the h–th stratum

(h = 1, 2, .. L); $\bar{y}_h = \frac{1}{n_h}\sum_{j=1}^{n_h} y_{hj}$ and $\bar{x}_h = \frac{1}{n_h}\sum_{j=1}^{n_{hl}} x_{hj}$ are the sample means of y and x respectively for the h–th stratum.

It is assumed that data in respect of auxiliary character x are available for all the units in the population.

The bias and mean square error of $\hat{Y}_{RS}$ to the first degree of approximation are respectively given by

$$B\,(\hat{Y}_{RS}) = \sum_{h=1}^{L} P_h \bar{Y}_h \frac{N_h - n_h}{N_h\, n_h} C^2_{hx}\,(1 - K_h) \tag{1.3}$$

and, $$MSE\,(\hat{Y}_{RS}) = \sum_{h=1}^{L} P^2_h \bar{Y}^2_h \frac{N_h - n_h}{N_h\, n_h} [C^2_{hy} + C^2_{hx}\,(1 - 2K_h)] \tag{1.4}$$

where, ρ_h is the correlation coefficient between y and x in the h–th stratum,

$$C_{hx} = (S_{hx}/\bar{X}_h), C_{hy} = (S_{hy}/\bar{Y}_h), S^2_{hx} = \frac{1}{(N_h - 1)} \sum_{j=1}^{N_h} (x_{hj} - \bar{X}_h)^2$$

$$S^2_{hy} = \frac{1}{(N_h - 1)} \sum_{j=1}^{N_h} (y_{hj} - \bar{Y}_h)^2 \text{ and } K_h = \rho_h\,(C_{hy}/C_{hx}).$$

We, now, define

$$\bar{x}^*_h = (N_h \bar{X}_h - n_h\, \bar{x}_h)/(N_h - n_h) \tag{1.5}$$

which is an unbiaesd estimator of $\bar{X}_h$, the population mean for the h–th stratum. The correlation between $\bar{y}_h$ and $\bar{x}^*_h$ is negative.

Thus, following Srivenkataramana (1980) we define a dual to ratio estimator for $\bar{Y}_h$ as

$$\bar{y}^{(1)}_{Rh*} = \bar{y}_h \, (\bar{x}^*_h/\bar{X}_h) \tag{1.6}$$

An interesting point with $\bar{y}^{*(1)}_{Rh}$ is that the exact expressions for bias and mean square error of $\bar{y}^{*}_{Rh}$ can be worked out which is not the case with $\bar{y}_{Rh}$ defined in (1.2).

Replacing $\bar{y}_{Rh}$ by $\bar{y}^{*}_{Rh}$ in (1.1) we obtain a dual to ratio estimator for $\bar{Y}$ as

$$\hat{Y}^{*}_{RS} = \sum_{h=1}^{L} P_h \, \bar{y}^{*(1)}_{Rh} \tag{1.7}$$

The exact bias and approximate MSE of $\hat{Y}^{*}_{RS}$ to the first degree of approximation are respectively given by

$$B(\hat{Y}^{*}_{RS}) = -(1/N) \sum_{h=1}^{L} K_h C^2_{hx} \tag{1.8}$$

and

$$MSE(\hat{Y}^{*}_{RS}) = \sum_{h=1}^{L} P^2_h \bar{Y}^2_h \frac{(N_h - n_h)}{n_h N_h} [C^2_{hy} + G_h C^2_{hx} (G_h - 2K_h)] \tag{1.9}$$

It is obvious from (1.8) that the estimator $(\hat{Y}^{*}_{RS})$ in (1.7) is a biased estimator. It is desired either to reduce or completely eliminate the bias. In this paper we have suggested a class of exactly unbiased dual to ratio estimator in stratified sampling and its properties are studied.

2. *The Class of Estimators*

Consider two estimators of $\bar{Y}_h$, the mean of h–th stratum in the population :

$$\bar{y}^{*(1)}_{Rh} = \bar{y}_h (\bar{x}^{*}_h / \bar{X}_h) \tag{2.1}$$

and

$$\bar{y}^{*(2)}_{Rh} = (1/\bar{X}_h) \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj} \, x^{*}_{hj} \tag{2.2}$$

where $x^{*}_{hj} = \{(N_h \bar{X}_h - n_h x_{hj})/(N_h - n_h)\}$

The biases of $\bar{y}^{*(1)}_{Rh}$ and $\bar{y}^{*(2)}_{Rh}$ are respectively given by

$$B(\bar{y}^{*(1)}_{Rh}) = -(1/N_h) \bar{Y}_h K_h C^2_{hx} \tag{2.3}$$

and

$$B(\bar{y}^{*(2)}_{Rh}) = -\{ n_h (N_h - 1)/N_h (N_h - n_h) \} \bar{Y}_h K_h C^2_{hx} \tag{2.4}$$

We, now, define a class of estimators for $\overline{Y}_h$, the mean of the h–th stratum in the population as

$$\hat{Y}_{Rh} = W_1 \bar{y}_h + W_2 \bar{y}^{(1)}_{Rh}* + W_3 \bar{y}^{(2)}_{Rh}* \tag{2.5}$$

where, W_i's $(i = 1, 2, 3)$ are suitably chosen constants such that

$$\sum_{i=1}^{3} W_i = 1 \tag{2.6}$$

The estimator $\hat{Y}_{Rh}$ would be unbiased if and only if

$$W_2 B(\bar{y}^{(1)}_{Rh}*) + W_3 B(\bar{y}^{(2)}_{Rh}*) = 0$$

or if
$$\delta_h W_2 + W_3 = 0 \tag{2.7}$$

where
$$\delta_h = (N_h - n_h)/n_h (N_h - 1) \tag{2.8}$$

From (2.6) and (2.7) we find that

$$W_1 = \{1 - (1 - \delta_h) W_2\} \tag{2.9}$$

Letting $W_2 = W_h$ (a constant) and putting $W_1 = [1 - (1 - \delta_h) W_h)$ and $W_3 = -\delta_h W_h$ in (2.5), we get a general class of unbiased dual to ratio type estimator for $\overline{Y}_h$ as

$$\hat{Y}^{(u)}_{Rh} = [\{1 - (1 - \delta_h) W_h\} \bar{y}_h + W_h \bar{y}^{(1)}_{Rh}* - W_h \delta_h \bar{y}^{(2)}_{Rh}*] \tag{2.10}$$

or,
$$\hat{Y}^{(u)}_{Rh} = [\{1 - (1 - \delta_h) W_h\} \bar{y}_h + W_h \bar{y}_h (\bar{x}_h*/\overline{X}_h) - W_h \frac{\delta h}{\overline{X}h} \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj} x^*_{hj}] \tag{2.11}$$

Hence, a class of unbiaed dual to ratio-type estimator for population mean $\overline{Y}$ based on all stratum mean estimators is given by

$$\hat{Y}^{(u)}_{RS} = \sum_{h=1}^{L} P_h \hat{Y}^{(u)}_{Rh} \tag{2.12}$$

or,
$$\hat{Y}^{(u)}_{RS} = \sum_{h=1}^{L} P_h \cdot \hat{Y}^{(u)}_{Rh} [\{1 - (1 - \delta_h) W_h\} \bar{y}_h + W_h \bar{y}_h (\bar{x}^*_h/\overline{X}_h)$$

$$- W_h (\delta_h/\overline{X}_h) \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj} x^*_{hj}] \tag{2.13}$$

Remark 2.1

(i) For $W_h = 0$, $\hat{Y}_{RS}^{(u)}$ gives the natural unbiaed estimator

$$\hat{Y}_{RS}^{(1)} = \sum_{h=1}^{L} P_h \bar{y}_h \tag{2.14}$$

while for $W_h = (1 - \delta_h)^{-1}$, it reduces to other unbiased estimator

$$\hat{Y}_{RS}^{(2)} = \sum_{h=1}^{L} P_h \left[\frac{N_h (n_h - 1)}{n_h (N_h - 1)} \bar{y}_h \frac{\bar{x}_h *}{\bar{X}_h} - \frac{(N_h - n_h)}{N_h (n_h - 1)} \frac{1}{\bar{X}_h} \frac{1}{nh} \sum_{j=1}^{n_h} y_{hj} x_{hj}^{*} \right] \tag{2.15}$$

(ii) For $W_h = 1$, $\hat{Y}_{RS}^{(u)}$ boils down to

$$\hat{Y}_{RS}^{(3)} = \sum_{h=1}^{L} P_h \left[\frac{N_h n_h)}{n_h (N_h - 1)} \{ \bar{y}_h - \frac{1}{n_h} \sum_{j=1} y_{hj} (x_{hj} * / \bar{X}_h) \} + \bar{y}_h \frac{\bar{x}_h *}{\bar{X}_h} \right] \tag{2.16}$$

Many other unbiased dual to ratio-type estimators for $\bar{Y}$ in stratified sampling can be had just by putting suitable values of W_h in (2.13).

3. *Optimum Estimator in (2.13)*

We have from (2.13) that

$$V(\hat{Y}_{RS}^{(u)}) = \sum_{h=1}^{L} P_h^2 \left[V(\bar{y}_h) + W_h^2 \{ V(\bar{y}_{Rh}^{(1)} *) + \delta_h^2 V(\bar{y}_{Rh}^{(2)} *) + (1 - \delta_h)^2 V(\bar{y}_h) \right.$$

$$- 2\delta_h \operatorname{CoV}(\bar{y}_{Rh}^{*}, \bar{y}_{Rh}^{(2)} *) - 2(1 - \delta_h) \operatorname{CoV}(\bar{y}_h, \bar{y}_{Rh}^{*(1)})$$

$$+ 2\delta_h (1 - \delta_h) \operatorname{CoV}(\bar{y}_h, \bar{y}_{Rh}^{(2)} *) \}$$

$$\left. + 2W_h \{ \operatorname{CoV}(\bar{y}_h, \bar{y}_{Rh}^{(1)} *) - \delta_h \operatorname{CoV}(\bar{y}_h, \bar{y}_{Rh}^{(2)} *) - (1 - \delta_h) V(\bar{y}_h) \} \right] \tag{3.1}$$

To the first degree of approximation, the variances and covariances involved in (3.1) are given as

$$V(\bar{y}_h) = \{ (N_h - n_h)/N_h n_h \} \bar{Y}_h^2 C_{hy}^2$$

$$V(\bar{y}_{Rh}^{(1)} *) = V(\bar{y}_{Rh}^{(2)} *) = \operatorname{Cov}(\bar{y}_{Rh}^{(1)} *, \bar{y}_{Rh}^{(2)} *)$$

$$= \frac{(N_h - n_h)}{N_h n_h} \bar{Y} h^2 [C_{hy}^2 + G_h C_{hx}^2 (G_h - 2K_h)]$$

$$\text{CoV}(\bar{y}_h, \bar{y}_{Rh}^{(1)}*) = \text{CoV}(\bar{y}_h, \bar{y}_{Rh}^{(2)}*) = \{(N_h - n_h)/N_h n_h\} \bar{Y}_h^2 [C_{hy}^2 - G_h C_{hx}^2 K_h] \tag{3.2}$$

Substitution of (3.2) in (3.1) yields the variance of $\hat{Y}_{RS}^{(u)}$ as

$$V(\hat{Y}_{RS}^{(u)}) = \sum_{h=1}^{L} P_h^2 \{(N_h - n_h)/n_h N_h\} \bar{Y}_h^2 [C_{hy}^2 + W_h (1-\delta_h) G_h C_{hx}^2 \{W_h (1-\delta_h) G_h - 2K_h\}] \tag{3.3}$$

which is minimised for

$$W_h = K_h/G_h (1 - \delta_h) \tag{3.4}$$

Thus, the resulting (minimum) variance of $\hat{Y}_{RS}^{(u)}$ is given by

$$\min. V(\hat{Y}_{RS}^{(u)}) = \sum_{h=1}^{L} P_h^2 \{(N_h - n_h)/N_h n_h\} \bar{Y}_h^2 C_{hy}^2 (1 - \rho_h^2) \tag{3.5}$$

Putting (3.4) in (2.13) we get the "*optimum estimator*" in the class (2.13) for $\bar{Y}$ as

$$\hat{Y}_{RS}^{(o)} = \sum_{h=1}^{L} P_h [\{1 - K_h/G_h)\} \bar{y}_h + \{k_h/G_h (1 - \delta_h))\} \bar{y}_h (\bar{x}_h*/\bar{X}_h) - \{\delta_h K_h/G_h (1 - \delta_h)\} \{1/n_h\} \sum_{j=1}^{n_h} y_{hj} (x_{hj}*/\bar{X}_h)\}] \tag{3.6}$$

with the variance

$$V(\hat{Y}_{RS}^{(o)}) = \min. V(\hat{Y}_{RS}^{(u)}) = \sum_{h=1}^{L} P_h^2 \{(N_h - n_h)/N_h n_h\} \bar{Y}_h^2 C_{hy}^2 (1 - \rho_h^2) \tag{3.7}$$

Remark 5.1

The variance of any estimator of the class $\hat{Y}_{RS}^{(u)}$ can be obtained from (3.3) to terms of order o (n^{-1}).

Remark 5.2

The "*optimum estimator*" $\hat{Y}_{RS}^{(o)}$ in (3.6) requires prior knowledge of $K_h = \rho_h (C_{hy}/C_{hx})$. Though exact value of K_h is rarely known, it is however, possible in repeated surveys or in studies based on multiphase sampling, where information on the same set of characteristics is collected over several occasions, to guess quite accurately the values of certain parameters. Thus, the value of K_h may be assessed to be used to obtain the feasible estimator.

4. *Theoretical Comparison*

The usual unbiased estimator of population mean $\bar{Y}$ in stratified sampling is given by

$$\hat{Y}_{RS}^{(1)} = \sum_{h=1}^{L} P_h \bar{y}_h$$

with variance

$$V(\hat{Y}_{RS}^{(1)}) = \sum_{h=1}^{L} P_h^2 \{(N_h - n_h)/n_h N_h\} \bar{Y}_h^2 C_{hy}^2 \tag{4.1}$$

It follows from (3.3) and (4.1) that the proposed class of unbiased dual to ratio-type estimator $\hat{Y}_{RS}^{(u)}$ is more efficient than usual unbiased estimator $\hat{Y}_{RS}^{(1)}$ if

$$V(\hat{Y}_{RS}^{(u)}) < V(\hat{Y}_{RS}^{(1)}) \tag{4.2}$$

or if

$$0 < W_h < \{2 K_h/G_h (1 - \delta_h)\}$$

It is to be noted from (1.4) and (3.3) that the estimator $\hat{Y}_{RS}^{(u)}$ is more efficient than usual ratio estimator $\hat{Y}_{RS}$ in stratified sampling if

$$V(\hat{Y}_{RS}^{(u)}) < MSE(\hat{Y}_{RS})$$

or if, either

$$\left.\begin{array}{l} \dfrac{1}{G_h(1-\delta_h)} < W_h < \dfrac{(2K_h - 1)}{G_h(1-\delta_h)} \\ \text{or} \\ \dfrac{(2K_h - 1)}{G_h(1-\delta_h)} < W_h < \dfrac{1}{G_h(1-\delta_h)} \end{array}\right] \tag{4.3}$$

Further, from (1.9) and (3.3) we note that the estimator $\hat{Y}_{RS}^{(u)}$ dominates over usual dual to ratio estimator in stratified sampling $\hat{Y}_{RS}^{*}$ if

$$V(\hat{Y}_{RS}^{(u)}) < MSE(\hat{Y}_{RS}^{*})$$

$$\left.\begin{array}{ll}\text{or if, either} & \dfrac{1}{(1-\delta_h)} < W_h < \dfrac{1}{(1-\delta_h)}\left(\dfrac{2K_h}{G_h}-1\right) \\ \text{or} & \dfrac{1}{(1-\delta_h)}\left(\dfrac{2K_h}{G_h}-1\right) < W_h < \dfrac{1}{(1-\delta_h)}\end{array}\right] \quad (4.4)$$

ACKNOWLEDGEMENT

The authors are grateful to the referee for his valuable comments leading to the improvement in the paper.

REFERENCE

[1] Srivenkataramana, T., 1980. A dual to ratio estimator in sample surveys. *Biometrika*, **67**, 199-204.

Jour. Ind. Soc. Ag. Statistics
47(2), 1995 : 176–183

The Effect of Survey Design on Regression Analysis : An Empirical Investigation

R.C. Agrawal and O.P. Kathuria*
National Bureau of Plant Genetic Resources, New Delhi-110012
(Received : September, 1993)

SUMMARY

An empirical investigation is carried out to study the effect of survey design on regression analysis. Under three different situations A, B and C, six different sample designs have been considered for the study. Under situation A, the complete population has been taken into consideration with X_1 as dependent, X_2 as independent variable and X_3 as design variable. In situation B and C, two phase sampling has been adopted; X_3 and X_1 have been used as design variables respectively. The bias of OLS estimator and mean square errors of other estimators have been compared under different sampling designs for the three situations.

Key Words : OLS estimator, Double sampling, Design variable, Regression analysis.

Introduction

In the complex survey design, the data is often analysed using regression techniques without further regard to the sample design (Nathan and Holt [4], Holt *et al.* [2]. The algebraic comparison of different estimators proposed by Nathan and Holt [4] in case of most of the sampling designs is difficult to put in practice. So, an empirical investigation is adopted for this comparison. For this purpose, data of a pilot sample survey for estimation of inland fishery resources and catch in a region of West Bengal, India is used.

2. *Description of the investigation*

The data from 1350 ponds obtained from a study conducted in India for developing sampling methodology for estimating the extent of area under ponds and catch of fish from them has been considered as the complete population (Kathuria et al [3]). For each pond, the observations on fish catch in Kg. (X_1), total quantity of fish seed used in Kg. (X_2) and the area of the pond in acres (X_3) have been taken from the survey. Variable X_3 is treated as the design variable.

The population of 1350 ponds has been divided into 5 strata, on the basis of the values of X_3 with stratum sizes 119, 516, 482, 153 and 80.

* Indian Agricultural Statistics Research Institute, New Delhi-110012