A class of almost unbiased estimators in systematic sampling

Sheela Misra

S. K. Pandey

Department of Statistics
Lucknow University
Lucknow 226 007
India

Abstract

In this paper, a general class of almost unbiased estimators is proposed for estimating population mean $\bar{Y}$ of the characteristic under study y when auxiliary information is available using systematic sampling plan. Explicit expressions for the variance of class of estimators is obtained to the first order of approximation. Minimum variance unbiased estimators (optimum estimators) in the class are also found. The class of estimators proposed by Kushwaha and Singh (1989) comes out to be a particular case of this class estimators.

Keywords : *Auxiliary, systematic sampling, ratio, product estimators, mean squared error, unbiased estimators, variance, bounded.*

1. Introduction

Many biased ratio type, product type transformed estimators obtained through linear combination of ratio or product and usual unbiased estimators are available for estimating the population mean when auxiliary information is available on auxiliary variable using simple random sampling. The classical ratio and product estimators under systematic sampling scheme were proposed by Swain (1964), and Shukla (1971) respectively and their properties were studied which was later generalized to a class of estimators by Kushwaha and Singh (1989). In the present paper a more general class of almost unbiased estimators is

Journal of Interdisciplinary Mathematics
Vol. 9 (2006), No. 2, pp. 363–371

proposed. The class of estimators proposed by Kushwaha and Singh (1988) and many others are the special cases of this class, using systematic sampling. The condition for unbiasedness and variance formula for this class of estimators are derived under large sample approximation.

2. Class of estimators

Assumes that a population consists of N units, serially numbered from 1 to N. Further assumes that $N = nk$ where n and k are integers. Now a systematic sample of size n is selected and both the characteristics x and y for each and every unit selected in the sample are observed. Here y and x denotes characteristics under study and auxiliary characteristic respectively. Let $(x_{ij} \cdot y_{ij})$ $i = 1, 2, \ldots, k,$ $j = 1, 2, \ldots, n$ denote the pair of the observed values of the jth unit in the sample. The systematic sample means are

$$\bar{y}_{sy} = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$$

$$\bar{x}_{sy} = \frac{1}{n} \sum_{j=1}^{n} x_{ij}$$

where $i = 1, 2, \ldots, k$.

The generalized estimator representing a class of estimators for estimating population mean $(\bar{Y})$ of the variable 'y' under study is proposed as

$$\bar{y}_{sp} = \bar{Y}_{sp} = \bar{y}_{sy} \cdot h\left(\frac{\bar{x}_{sy}}{\bar{X}}\right) = \bar{y}h(u) \tag{2.1}$$

where $\bar{X}$ is the population mean of the auxiliary variable x and $u = \frac{\bar{x}_{sy}}{\bar{X}} \cdot h(u)$, such that $h(1) = 1$, is a function of u, satisfying the following conditions:

i. whatever be the sample chosen, u assumes values in bounded closed interval I of the real line containing the point unity.

ii. In I, the function $h(u)$ is continuous and bounded.

iii. The first, second and third partial derivatives of $h(u)$ exist and are continuous and bounded in I.

Expanding $h(u)$ about $u = 1$ in the third order Taylor's series, we have from (1.1)

$$\bar{y}_{Sp} = \bar{y}_{sy} h(1) + h(1) \cdot (u-1) + \frac{(u-1)^2}{2} h''(1) + \frac{(u-1)^3}{3} h''(u^*) \quad (2.2)$$

where $u^* = 1 + \theta(u-1)$, $0 < \theta < 1$, and θ may depend on $u \cdot h'(1)$, $h''(1)$, $h'''(u^*)$ denote the first, second and third partial derivatives of $h(u)$ at the point $u = 1$, 1 and u^*, respectively.

Now take $n = gm$ and split the systematic sample of size following class of estimators based on sub-samples

$$\bar{y}_{spt} = \frac{1}{g} \sum_{t=1}^{g} \bar{y}_t h(u_t) \quad (2.3)$$

where $u_t = \dfrac{\bar{x}_{syt}}{\bar{X}_t}$ and $h(1) = 1$.

3. Biases of $\bar{y}_{Sp}$ and $\bar{y}_{spt}$

The expressions for biases of $\bar{y}_{sp}$ and $\bar{y}_{spt}$ population correction for large nk to the first order of approximation are obtained as

$$\text{Bias}(\bar{y}_{sp}) = \frac{\bar{Y}}{n} \{1 + (n-1)\rho_w\} C_x^2 \left\{ h'(1) \frac{\rho_{yx} \cdot C_y}{C_x} + \frac{h''(1)}{2} \right\} \quad (3.1)$$

and

$$\text{Bias}(\bar{y}_{spt}) = \frac{\bar{Y}}{n} \{g + (n-g)\rho_w\} C_x^2 \left\{ h'(1) \frac{\rho_{yx} \cdot C_y}{C_x} + \frac{h''(1)}{2} \right\} \quad (3.2)$$

where ρ_{yx} is the correlation coefficient between y and x variables in the population, and ρ_w is the intra class correlation coefficient, assumed to be same for both the characteristic x and y. It has also been assumed to be known. C_x and C_y are the coefficient of variation for the characteristic x and y, respectively.

4. Weighted class of estimator for $\bar{Y}$

A weighted class of estimators for $\bar{Y}$ is proposed as

$$T_{Sp} = \ell_1 \bar{y}_{sy} + \ell_3 \bar{y}_{Sp} + \ell_2 \bar{y}_{spt} \quad (4.1)$$

where $\ell_1 + \ell_2 + \ell_3 = 1$, and are suitably chosen weights given to different estimators.

Theorem 4.1. *The weighted class of estimators* T_{Sp} *proposed in* (4.1) *is unbiased for population mean* $\bar{Y}$ *if and only if*

$$p\ell_2 + \ell_3 = 0$$

where

$$p = \frac{g + (n-g)\rho_w}{1 + (n-1)\rho_w}. \tag{4.2}$$

Taking $\ell_1 = \ell$, $\ell_2 = \ell'$ and $\ell_3 = 1 - \ell - \ell'$, the unbiasedness condition reduces to

$$\ell' = -\left(\frac{1-\ell}{p-1}\right)$$

where ℓ and ℓ' are constants to be chosen suitable. Thus a general class of almost unbiased estimators may be obtained as

$$T_{Spv} = \bar{y}_{sy} + \left(\frac{1-\ell}{p-1}\right) p \cdot \bar{y}_{sp} - \left(\frac{1-\ell}{p-1}\right) \bar{y}_{spt}. \tag{4.3}$$

5. Properties of the class

The variance of the proposed class T_{Sp} given in (4.1) is given by

$$\begin{aligned} \operatorname{Var}(T_{Sp}) = \; & \ell_1^2 \cdot \operatorname{Var}(\bar{y}_{sy}) + \ell_2^2 \cdot \operatorname{Var}(\bar{y}_{sp}) + \ell_3^2 \cdot \operatorname{Var}(\bar{y}_{spt}) \\ & + 2\ell_1\ell_2 \operatorname{Cov}(\bar{y}_{sy}, \bar{y}_{Sp}) + 2\ell_2\ell_3 \operatorname{Cov}(\bar{y}_{sp}, \bar{y}_{spt}) \\ & + 2\ell_1\ell_3 \operatorname{Cov}(\bar{y}_{sy}, \bar{y}_{spt}). \end{aligned} \tag{5.1}$$

To the first order approximation the variance and covariance expressions for various estimators and class of estimators in (5.1) are cited and proved in the lemma (5.1).

Lemma 5.1.

$$\begin{aligned} \operatorname{Var}(\bar{y}_{Sy}) &= \frac{\bar{Y}^2}{n}\{1 + (n-1)\rho_w\} c_y^2 \\ \operatorname{Var}(\bar{y}_{Spt}) &= \operatorname{Var}(\bar{y}_{Sp}) = \operatorname{Cov}(\bar{y}_{Sp}, \bar{y}_{Spt}) \\ &= \frac{\bar{Y}^2}{n}\{1 + (n-1)\rho_w\}\left\{C_y^2 + h'(1)C_x^2 \cdot \left(h'(1) + \frac{2\rho_{yx}C_y}{C_x}\right)\right\} \\ \operatorname{Cov}(\bar{y}_{Sy}, \bar{y}_{Spt}) &= \operatorname{Cov}(\bar{y}_{Sy}, \bar{y}_{sp}) \\ &= \frac{\bar{Y}_2}{n}\{1 + (n-1)\rho_w\}\left\{C_y^2 + h'(1)\frac{2\rho_{yx}C_y}{C_x}\right\}. \end{aligned} \tag{5.2}$$

Proof. It may easily be verified that

$$V(\bar{y}_{spt}) = V(\bar{y}_{sp}) = \text{Cov}(\bar{y}_{spt}, \bar{y}_{sp})$$

and

$$\text{Cov}(\bar{y}_{sy}, \bar{y}_{spt}) = \text{Cov}(\bar{y}_{sy}, \bar{y}_{sp})\,.$$

To find the expressions for $\text{Cov}(\bar{y}_{Spt} > \bar{y}_{sp})$ and $\text{Cov}(\bar{y}_{sy}, \bar{y}_{sp})$ under large sample approximation, let

$$\bar{y}_{sy} = \bar{Y}(1+e_0) \qquad \text{or} \qquad e_0 = \frac{\bar{y}_{sy} - \bar{Y}}{\bar{Y}}\,,$$

$$\bar{x}_{sy} = \bar{X}(1+e_1) \qquad \text{or} \qquad e_1 = \frac{\bar{x}_{sy} - \bar{Y}}{\bar{x}}\,.$$

From (1 .2) we have

$$\bar{y}_{Sp} - \bar{Y} = \bar{Y}\left\{h'(1)e_1 + e_0 + \frac{h''(1)}{2}e_1^2 + e_0 e_1 h'(1)\right\}. \tag{5.3}$$

To the first order of approximation.

Again let

$$\begin{aligned} \bar{y}_{syt} &= \bar{Y}(1+e_0')\,, \\ \bar{x}_{syt} &= \bar{X}(1+e_1')\,. \end{aligned}$$

Expanding $h(u_t)$ about $u_t = 1$ in the third order Taylor's series in (2.2) and ignoring terms of power higher than two in e_i's, we obtain

$$\bar{y}_{Spt} - \bar{Y} = \bar{Y}\left\{h'(1)e_1' + e_0' + \frac{h''(1)e_1'^2}{2} + h'(1)e_0'e_1'\right\}. \tag{5.4}$$

Using (5.3) and (5.4), we have

$$\begin{aligned} \text{Cov}(\bar{y}_{Sp}, \bar{y}_{Spt}) &= E(\bar{y}_{Sp} - \bar{Y})(\bar{y}_{spt} - \bar{Y}) \\ &= \bar{Y}^2 E[h'(1)^2 e_1 e_1' + e_0 e_0' + h'(1)(e_1 e_0' + e_0 e_1')] \\ &\qquad \text{(to the first order of approximation)} \\ &= \bar{Y}^2 E\left[(h'(1))^2 e_1 E_2\left(\frac{e_1'}{t}\right) + e_0 E_2\left(\frac{e_0'}{t}\right)\right. \\ &\qquad \left. + h'(1)\left\{e_1 E_2\left(\frac{e_0'}{t}\right) e_0 E_2\left(\frac{e_1'}{t}\right)\right\}\right] \end{aligned} \tag{5.5}$$

where $E_2\left(\frac{e_0'}{t}\right) = \frac{E_2\left(\frac{\bar{y}_{syt}}{t}\right) - \bar{Y}}{\bar{Y}}$ is the conditional expectation for a given tth $(t = 1, 2, \ldots, 9)$ split and E_1 is the expectation on ith $(i = 1, 2, \ldots, k)$ systematic sample of size n. Here, it should be noted that $\bar{y}_{syt}$ is a systematic sample mean of size m drawn from a population of size n and therefore,

$$E\left(\frac{\bar{y}_{syt}}{t}\right) = \bar{y}_{sy}$$

or

$$E_2\left(\frac{e_0'}{t}\right) = e_0$$

and similarly

$$E_2\left(\frac{e_1'}{t}\right) = e_1 .$$

Hence from (5.5), we have

$$\begin{aligned}
\mathrm{Cov}(\bar{y}_{sp}, \bar{y}_{Spt}) &= \bar{Y}^2 E_1\{(h_1'(1))^2 e_1^2 + e_0^2 + h'(1)\cdot(e_1e_0 + e_0e_1)\} \\
&= \bar{Y}_2\left[(h'(1))^2 \frac{\mathrm{Var}(\bar{x}_{sy})}{\bar{X}^2} + \frac{\mathrm{Var}(\bar{y}_{sy})}{\bar{Y}_2}\right. \\
&\qquad \left. + h'(1)\cdot\frac{2\mathrm{Cov}(\bar{x}_{sy}, \bar{y}_{Sy})}{\bar{X}\bar{Y}}\right] \\
&= \frac{\bar{Y}^2}{n}\{1 + (n-1)\rho_w\}\left[C_y^2 h'(1) C_x^2 (h'(1)) + \frac{2\rho_{yx}C_y}{C_x}\right]. \qquad (5.6)
\end{aligned}$$

Similarly for $\mathrm{Cov}(\bar{y}_{sy}, \bar{y}_{sp})$ we have,

$$\begin{aligned}
\mathrm{Cov}(\bar{y}_{sy}, \bar{y}_{sp}) &= E(\bar{y}_{sy} - \bar{Y})(\bar{y}_{Sp} - \bar{Y}) \\
&= \bar{Y}^E\left[e_0 \cdot \left\{h'(1)e_1 + e_0 + \frac{h'(1)}{2}e_1^2 + h'(1)e_0e_1\right\}\right] \\
&\qquad \text{(Using (2.2) to the 1st order of approximation)} \\
&= \bar{Y}^2 E_1\{h'(1)e_0e_1 + e_0^1\} \\
&= \bar{Y}^2\left\{\frac{\mathrm{Var}(\bar{y}_{sy})}{\bar{Y}^2} + h'(1)\cdot\frac{\mathrm{Cov}(\bar{y}_{sy}, \bar{x}_{sy})}{\bar{X}\bar{Y}}\right\} \\
&= \frac{\bar{Y}^2}{2}\{(1 + (n-1)\rho_w\}[C_y^2 - h'(1)\rho_{yx}C_yC_x] . \qquad (5.7)
\end{aligned}$$

Substituting the result (5.2) and $\ell_1 = \ell$, $\ell_2 = \ell'$ and $\ell_3 = 1 - \ell - \ell'$ in (4.1), we obtain the variance formula for the class of estimators T_{Sp} given as

$$\begin{aligned} \mathrm{Var}(T_{sp}) &= \frac{\bar{Y}^2}{n}\{1+(n-1)\rho_w\} \\ &\quad \times \left[C_y^2 + (1-\ell)h'(1)C_x^2\left\{(1-\ell)h'(1) + 2\rho_{yx}\frac{C_y}{C_x}\right\}\right] \end{aligned} \tag{5.8}$$

The variance of T_{Sp} in (4.8) will be min for

$$\ell = 1 + \frac{\rho_{yx}C_y}{(h'(1))C_x} = \ell^* \quad \text{(say)}. \tag{5.9}$$

Thus, the minimum variance of T_{Sp} is given by

$$\begin{aligned} \text{Min } \mathrm{Var}(T_{Sp}) &= \frac{\bar{Y}^2}{n}\{1+(n-1)\rho_w\}[C_y^2 - \rho_{yx}^2 C_y^2] \\ &= \frac{\bar{Y}^2}{n}\{1+(n-1)\rho'_w\}(1-\rho_{yx}^2)C_y^2 \end{aligned} \tag{5.10}$$

which is equivalent to the approximate variance of usual biased linear regression estimator $\bar{y}_{er}$ in systematic sampling given as

$$\bar{y}_{er} = \bar{y}_{Sy} + b_{yx}(\bar{X} - \bar{x}_{sy})$$

where b_{yx} is the sample regression coefficient of y on x.

Substituting g,

$$\left.\begin{aligned} &\ell_1 = \ell^* = 1 + \left(\frac{\rho_{yx}\cdot C_y}{h'(1)C_x}\right); \\ &\ell_2 = \ell'^* = \frac{\rho_{yx}\cdot C_y}{C_x h'(1)(p-1)} \\ &\text{and} \\ &\ell_3 = 1 - \ell^* - \ell'^* = -\frac{\rho_{yx}\cdot C_y}{h'(1)C_x}\left(\frac{p}{p-1}\right) \end{aligned}\right\} \tag{5.11}$$

in (4.1), we obtain an optimum estimator in the class T_{Sp} given as

$$\begin{aligned} T_{Spo} &= \left(1 + \frac{\rho_{yx}C_y}{h'(1)C_x}\right)\cdot \bar{y}_{Sy} + \frac{\rho_{yx}C_y}{C_x h'(1)(p-1)}\cdot \bar{y}_{spt} \\ &\quad - \frac{\rho_{yx}\cdot C_y}{h'(1)C_x}\left(\frac{p}{p-1}\right)\cdot \bar{y}_{sp} \end{aligned} \tag{5.12}$$

6. Concluding remarks

Take function $h(u)$ such that

$$H(u) = u\,; \qquad u = \frac{\bar{x}_{sy}}{\bar{X}}\,.$$

First and second partial derivatives of $h(u)$ with respect to u at $u = 1$, will be

$$h'(1) = 1$$

and

$$h''(1) = 0$$

Putting these values in (1.1) and (1.2) we have

$$\bar{y}_{Sp} = \frac{\bar{y}_{sy} \cdot \bar{x}_{sy}}{\bar{X}} \tag{6.1}$$

and

$$\bar{y}_{spt} = \frac{1}{g} \sum_{t=1}^{g} \bar{y}_{syt} \cdot \frac{\bar{x}_{syt}}{\bar{X}}\,. \tag{6.2}$$

$\bar{y}_{Sp}$ given by (6.1) comes out to be estimator proposed by Shukla (1971). Its variance can be found from (5.2).

Substituting (6.1) and (6.2) in (4.1) we get the class of product type estimators proposed by Kushwaha and Singh (1989). Now, let us take

$$h(u) = u^{-1}$$

then first and second partial derivatives of $h(u)$ with respect to 'u' at point $u = 1$ will be

$$h'(1) = -1\,, \qquad h''(1) = 2\,.$$

Putting these values in (1.1) and (1.2) we have

$$\bar{y}_{Sp} = \bar{y}_{sy} \left(\frac{\bar{X}}{\bar{x}_{sy}} \right) \tag{6.3}$$

and

$$\bar{y}_{spt} = \frac{1}{g} \sum_{t=1}^{g} \left(\frac{\bar{y}_{syt}}{\bar{x}_{syt}} \right) \cdot \bar{X}\,. \tag{6.4}$$

Substituting (6.3) and (6.4) in (4.1) we obtain class of Ratio type estimators proposed by Kushwaha and Singh (1989).

References

[1] K. S. Kushwaha and H. P. Singh (1989), Almost unbiased ratio and product estimators in systematic sampling, *Journal of the Indian Society of Agricultural Statistics*, Vol. XLI (2), pp. 193–205.

[2] P. V. Sukhatme and B. V. Sukhatme (1970), *Sampling Theory of Surveys with Applications*, IOWA State University Press, Ames, IOWA, USA.

[3] N. D. Shukla (1971), Systematic sampling and product method of estimation, in *Proceeding of All India Seminar on Demography and Statistics*, BHU, Varanasi, India.