ALMOST UNBIASED RATIO AND PRODUCT TYPE ESTIMATOR OF FINITE POPULATION VARIANCE USING THE KNOWLEDGE OF KURTOSIS OF AN AUXILIARY VARIABLE IN SAMPLE SURVEYS

Rajesh Singh, Pankaj Chauhan, and Nirmala Sawan,

School of Statistics, DAVV, Indore (M.P.), India

(rsinghstat@yahoo.com)

Florentin Smarandache
Chair of Department of Mathematics, University of New Mexico, Gallup, USA
(smarand@unm.edu)

Abstract

It is well recognized that the use of auxiliary information in sample survey design results in efficient estimators of population parameters under some realistic conditions. Out of many ratio, product and regression methods of estimation are good examples in this context. Using the knowledge of kurtosis of an auxiliary variable Upadhyaya and Singh (1999) has suggested an estimator for population variance. In this paper, following the approach of Singh and Singh (1993), we have suggested almost unbiased ratio and product-type estimators for population variance.

1. Introduction

Let $U = (U_1, U_2, ..., U_N)$ denote a population of N units from which a simple random sample without replacement (SRSWOR) of size n is to be drawn. Further let y and x denote the study and the auxiliary variables respectively. The problem is to estimate the parameter

$$S_y^2 = \frac{N}{N-1}\sigma_y^2 \tag{1.1}$$

with $\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2$ of the study variate y when the parameter

$$S_x^2 = \frac{N}{N-1}\sigma_x^2 \tag{1.2}$$

with $\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$ of the auxiliary variate x is known,

where $\overline{Y} = \sum_{i=1}^{N}\frac{y_i}{N}$ and $\overline{X} = \sum_{i=1}^{N}\frac{x_i}{N}$; are the population means of y and x respectively.

The conventional unbiased estimator of S_y^2 is defined by

$$s_y^2 = \frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)^2}{(n-1)} \tag{1.3}$$

where $\overline{y} = \sum_{i=1}^{n}\frac{y_i}{n}$ is the sample mean of y.

Using information on S_x^2, Isaki (1983) proposed a ratio estimator for S_y^2 as

$$t_1 = s_y^2\frac{S_x^2}{s_x^2} \tag{1.4}$$

where $s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ is unbiased estimator of S_x^2.

In many survey situations the values of the auxiliary variable x may be available for each unit in the population. Thus the value of the kurtosis $\beta_2(x)$ of the auxiliary variable x is known. Using information on both S_x^2 and $\beta_2(x)$ Upadhyay and Singh (1999) suggested a ratio type estimator for S_y^2 as

$$t_2 = s_y^2 \left[\frac{S_x^2 + \beta_2(x)}{s_x^2 + \beta_2(x)} \right] \qquad (1.5)$$

For simplicity suppose that the population size N is large enough relative to the sample size n and assume that the finite population correction (fpc) term can be ignored. Up to the first order of approximation, the variance of s_y^2, and t_1 and bias and variances of t_2 (ignoring fpc term) are respectively given by

$$\mathrm{var}(s_y^2) = \frac{S_y^4}{n}\{\beta_2(y) - 1\} \qquad (1.6)$$

$$\mathrm{var}(t_1) = \frac{S_y^4}{n}[\{\beta_2(y) - 1\} + \{\beta_2(x) - 1\}(1 - 2C)] \qquad (1.7)$$

$$B(t_2) = \frac{S_y^2}{n}[\{\beta_2(x) - 1\}\theta(\theta - C)] \qquad (1.8)$$

$$\mathrm{var}(t_2) = \frac{S_y^4}{n}[\{\beta_2(y) - 1\} + \theta\{\beta_2(x) - 1\}(\theta - 2C)] \qquad (1.9)$$

where $\theta = \frac{S_x^2}{S_x^2 + \beta_2(x)}$; $\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}$; $\beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}$; $h = \frac{\mu_{22}}{(\mu_{20}.\mu_{02})}$; $C = \frac{(h-1)}{\beta_2(x) - 1}$ and

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

From (1.8), we see that the estimator t_2 suggested by Upadhyay and Singh (1999) is a biased estimator. In some application bias is disadvantageous. This led authors to suggest almost unbiased estimators of S_y^2.

2. A class of ratio-type estimators

Consider $t_{Ri} = s_y^2 \left(\frac{S_x^2 + \beta_2(x)}{s_x^2 + \beta_2(x)} \right)^i$ such that $t_{Ri} \in R$, for $i = 1,2,3$; where R denotes the set of all possible ratio-type estimators for estimating the population variance S_y^2. We define a class of ratio-type estimators for S_y^2 as –

$$t_r = \sum_{i=1}^{3} w_i t_{Ri} \in R, \tag{2.1}$$

where $\sum_{i=1}^{3} w_i = 1$ and w_i are real numbers. (2.2)

For simplicity we assume that the population size N is large enough so that the fpc terms are ignored. We write

$$s_y^2 = S_y^2(1+e_0), s_x^2 = S_x^2(1+e_1)$$

such that E (e_0)=E (e_1)=0.

Noting that for large N, $\frac{1}{N} \cong 0$ and $\frac{n}{N} \cong 0$, and thus to the first degree of approximation,

$$E(e_0^2) = \frac{\beta_2(y)-1}{n},\ E(e_1^2) = \frac{\beta_2(x)-1}{n},\ E(e_0 e_1) = \frac{(h-1)}{n} = \frac{[\beta_2(x)-1]C}{n}.$$

Expressing (2.1) in terms of e's we have

$$t_r = S_y^2(1+e_0)\sum_{i=1}^{3} a_i (1+\theta e_1)^{-i} \tag{2.3}$$

Assume that $|\theta e_1| < 1$ so that $(1+\theta e_1)^i$ is expandable. Thus expanding the right hand side of the above expression (2.3) and retaining terms up to second power of e's , we have

$$t_r = S_y^2\left[1 + e_0 - \sum_{i=1}^{3} a_i i\left(\theta e_1 + \theta e_0 e_1 - \left(\frac{i+1}{2}\right)\theta^2 e_1^2\right)\right]$$

or

$$t_r - S_y^2 = S_y^2\left[e_0 - \sum_{i=1}^{3} a_i i\left(\theta e_1 + \theta e_0 e_1 - \left(\frac{i+1}{2}\right)\theta^2 e_1^2\right)\right] \quad (2.4)$$

Taking expectation of both sides of (2.3) we get the bias of t_r , to the first degree of approximation, as

$$B(t_r) = \frac{S_y^2}{2n}\left[\{\beta_2(x)-1\}\sum_{i=1}^{3} ia_i\theta(\theta i - 2C + \theta)\right] \quad (2.5)$$

Squaring both sides of (2.4), neglecting terms involving power of e's greater than two and then taking expectation of both sides, we get the mean-squared error of t_r to the first degree of approximation, as

$$MSE(t_r) = \frac{S_y^4}{n}\left[\{\beta_2(y)-1\} + R_1\{\theta\beta_2(x)-1\}\{\theta R_1 - 2C\}\right] \quad (2.6)$$

where $R_1 = \sum_{i=1}^{3} i.w_i$ (2.7)

Minimizing the MSE of t_r in (2.7) with respect to R_1 we get the optimum value of R_1 as

$$R_1 = \frac{C}{\theta} \quad (2.8)$$

Thus the minimum MSE of t_r is given by

$$\min.MSE(t_r) = \frac{S_y^4}{n}\left[\{\beta_2(y)-1\} - \{\beta_2(x)-1\}C^2\right]$$

$$= \frac{S_y^4}{n}\left[\{\beta_2(y)-1\}(1-\rho_1^2)\right] \quad (2.9)$$

where $\rho_1 = \frac{(h-1)}{\sqrt{\{\beta_2(x)-1\}\{\beta_2(y)-1\}}}$ is the correlation coefficient between $(y-\bar{Y})^2$ and $(x-\bar{X})^2$.

From (2.2), (2.7) and (2.8) we have

$$\sum_{i=1}^{3} w_i = 1 \qquad (2.10)$$

$$\text{and } \sum_{i=1}^{3} i w_i = \frac{C}{\theta} = \frac{\rho_1}{\theta}\left\{\frac{\beta_2(y)-1}{\beta_2(x)-1}\right\}^{\frac{1}{2}} \qquad (2.11)$$

From (2.10) and (2.11) we have three unknown to be determined from two equations only. It is therefore, not possible to find a unique value of the constants $w_i's(i=1,2,3)$. Thus in order to get the unique values of the constants $w_i's(i=1,2,3)$, we shall impose a linear constraint as

$$B(t_r) = 0 \qquad (2.12)$$

which follows from (2.5) that

$$(\theta - C)a_1 + (3\theta - 2C)a_2 + (6\theta - 3C)a_3 = 0 \qquad (2.13)$$

Equation (2.10), (2.11) and (2.13) can be written in the matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ (\theta - C) & (3\theta - 2C) & (6\theta - 3C) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 \\ C / \theta \\ 0 \end{bmatrix} \qquad (2.14)$$

Using (2.14) we get the unique values of $w_i's(i=1,2,3)$ as

$$\left.\begin{aligned} w_1 &= \frac{1}{\theta^2}\left[3\theta^2 - 3\theta C + C^2\right] \\ w_2 &= \frac{1}{\theta^2}\left[-3\theta^2 + 5\theta C - 2C^2\right] \\ w_3 &= \frac{1}{\theta^2}\left[\theta^2 - 2\theta C + C^2\right] \end{aligned}\right\} \qquad (2.15)$$

Use of these $w_i's(i=1,2,3)$ remove the bias up to terms of order $o(n^{-1})$ at (2.1). Substitution of (2.14) in (2.1) yields the almost unbiased optimum ratio-type estimator of the population variance S_y^2.

3. A class of product-type estimators

Consider $t_{Pi} = s_y^2\left[\frac{s_x^2+\beta_2(x)}{S_x^2+\beta_2(x)}\right]^i$ such that $t_{Pi} \in P$, for $i=1,2,3$; where P denotes the set of all possible product-type estimators for estimating the population variance S_y^2.

We define a class of product-type estimators for S_y^2 as –

$$t_P = \sum_{i=1}^{3} k_i t_{Pi} \in P \,, \tag{3.1}$$

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

$\sum_{i=1}^{3} k_i = 1$ and k_i are real numbers.

Proceeding as in previous section, we get

$$B(t_P) = \frac{S_y^2}{2n}\left[\{\beta_2(x)-1\}\sum_{i=1}^{3} i a_i \theta(\theta i + 2C - \theta)\right] \tag{3.2}$$

$$MSE(t_P) = \frac{S_y^4}{n}\left[\{\beta_2(y)-1\} + R_2\theta\{(\beta_2(x)-1)\}(\theta R_2 + 2C)\right] \tag{3.3}$$

where, $R_2 = \sum_{i=1}^{3} i k_i$ (3.4)

Minimizing the MSE of t_P in (3.4) with respect to R_2, we get the optimum value of R_2 as

$$R_2 = -\frac{C}{\theta} \tag{3.5}$$

Thus the minimum MSE of t_P is given by

$$\min.MSE(t_P) = \frac{S_y^4}{n}\{\beta_2(y) - 1\}(1 - \rho_1^2) \tag{3.7}$$

which is same as that of minimum MSE of t_r at (2.9).

Following the approach of previous section, we get

$$\left.\begin{aligned} k_1 &= \frac{1}{\theta^2}\left[3\theta^2 + 2\theta C + C^2\right] \\ k_2 &= -\frac{1}{\theta^2}\left[3\theta^2 + 3\theta C + 2C^2\right] \\ k_3 &= \frac{1}{\theta^2}\left[\theta^2 + \theta C + C^2\right] \end{aligned}\right\} \tag{3.8}$$

Use of these k_i's (i=1,2,3) removes the bias up to terms of order O (n^{-1}) at (3.1).

4. Empirical Study

The data for the empirical study are taken from two natural population data sets considered by Das (1988) and Ahmed et.al. (2003).

Population I – Das (1988)

The variables and the required parameters are:

X: number of agricultural labourers for 1961.

Y: number of agricultural labourers for 1971.

$\beta_2(x) = 38.8898$, $\beta_2(y) = 25.8969$, h=26.8142, $S_x^2 = 1654.44$.

Population II – Ahmed et.al. (2003)

The variables and the required parameters are:

X: number of households

Y: number of literate persons

$\beta_2(x) = 8.05448,\ \beta_2(y) = 10.90334,\ S_x^2 = 11838.85$, h=7.31399.

In table 4.1 the values of scalars w_i's (i=1,2,3) and k_i's (i=1,2,3) are listed.

Table 4.1: Values of scalars w_i's and k_i's (i=1,2,3)

Scalars	Population		Scalars	Population	
	I	**II**		**I**	**II**
w_1	1.3942	1.1154	k_1	4.8811	5.5933
w_2	-0.4858	-0.1261	k_2	-6.0647	-7.2910
w_3	0.0916	0.0109	k_3	2.1837	2.6978

Using these values of w_i's and k_i's (i=1,2,3) given in table 4.1,one can reduce the bias to the order $O(n^{-1})$ respectively, in the estimators t_r and t_p at (2.1) and (3.1).

In table 4.2 percent relative efficiency (PRE) of s_y^2, t_1, t_2, t_r (in optimum case) and t_p (in optimum case) are computed with respect to s_y^2.

Table 4.2: PRE of different estimators of S_y^2 with respect to s_y^2

Estimators	PRE (., S_y^2)	
	Population I	**Population II**
s_y^2	100	100

t_1	223.14	228.70
t_2	235.19	228.76
t_r (optimum)	305.66	232.90
t_p (optimum)	305.66	232.90

Table 4.2 clearly shows that the suggested estimators t_r and t_p in their optimum case are better than the usual unbiased estimator s_y^2, Isaki's (1983) estimator t_1 and Upadhayaya and Singh (1999) estimator t_2.

References

Ahmed, M.S., Abu Dayyeh, W. and Hurairah, A. A. O. (2003): Some estimators for finite population variance under two-phase sampling. Statistics in Transition, 6, (1), 143-150.

Das, A.K. (1988): Contributions to the theory of sampling strategies based on auxiliary information. Ph.D thesis submitted to BCKV, Mohanpur, Nadia, and West Bengal, India.

Isaki, C. T. (1983): Variance estimation using auxiliary information. Journal of American Statistical Association.

Singh, S. and Singh, R. (1993): A new method: Almost Separation of bias precipitates in sample surveys. Journal of Indian Statistical Association, 31,99-105.

Upadhyaya, L.N. and Singh, H. P. (1999): An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. Vikram Mathematical Journal, 19, 14-17.