Viplav Kumar Singh, Rajesh Singh

Department of Statistics, Banaras Hindu University

Varanasi-221005, India

Florentin Smarandache

University of New Mexico,

Gallup, USA

Some Improved Estimators for Population Variance Using Two Auxiliary Variables in Double Sampling

Published in:
Rajesh Singh, Florentin Smarandache (Editors)
ON IMPROVEMENT IN ESTIMATING POPULATION PARAMETER(S) USING AUXILIARY INFORMATION
Educational Publishing (Columbus) & Journal of Matter Regularity (Beijing), USA - China, 2013
ISBN: 978-1-59973-230-5
pp. 54 - 64

Abstract

In this article we have proposed an efficient generalised class of estimator using two auxiliary variables for estimating unknown population variance S_y^2 of study variable y .We have also extended our problem to the case of two phase sampling. In support of theoretical results we have included an empirical study.

1. Introduction

Use of auxiliary information improves the precision of the estimate of parameter .Out of many ratio and product methods of estimation are good example in this context. We can use ratio method of estimation when correlation coefficient between auxiliary and study variate is positive (high), on the other hand we use product method of estimation when correlation coefficient between auxiliary and study variate is highly negative.

Variations are present everywhere in our day-to-day life. An agriculturist needs an adequate understanding of the variations in climatic factors especially from place to place (or time to time) to be able to plan on when, how and where to plant his crop. The problem of estimation of finite population variance S_y^2, of the study variable y was discussed by Isaki (1983), Singh and Singh (2001, 2002, 2003), Singh et al. (2008), Grover (2010), and Singh et al. (2011).

Let x and z are auxiliary variates having values (x_i, z_i) and y is the study variate having values (y_i) respectively. Let $V_i (i = 1,2,.......N)$ is the population having N units such that y is positively correlated x and negatively correlated with z. To estimate S_y^2, we assume that S_x^2 and S_z^2 are known, where

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{Y})^2, \quad S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2 \text{ and } S_z^2 = \frac{1}{N}\sum_{i=1}^{N}(z_i - \overline{Z})^2.$$

Assume that N is large so that the finite population correction terms are ignored. A sample of size n is drawn from the population V using simple random sample without replacement.

Usual unbiased estimator of population variance S_y^2 is s_y^2, where, $s_y^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \overline{y})^2$.

Up to the first order of approximation, variance of s_y^2 is given by

$$\text{var}(s_y^2) = \frac{S_y^4}{n}\partial_{400}^* \tag{1.1}$$

where, $\partial_{400}^* = \partial_{400} - 1$, $\quad \partial_{pqr} = \frac{\mu_{pqr}}{\mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2}}$, and

$\mu_{pqr} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{Y})^p (x_i - \overline{X})^q (z_i - Z)^r$; p, q, r being the non-negative integers.

2. Existing Estimators

Let $s_y^2 = S_y^2(1+e_0), s_x^2 = S_x^2(1+e_1)$ and $s_z^2 = S_z^2(1+e_2)$

where, $s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2, s_z^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(z_i - \overline{z})^2$

and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i, \quad \overline{z} = \frac{1}{n}\sum_{i=1}^{n}z_i$.

Also, let

$E(e_1) = E(e_2) = 0$

$$E(e_0^2)=\frac{\partial_{400}^*}{n}, E(e_1^2)=\frac{\partial_{040}^*}{n} \text{ and } E(e_2^2)=\frac{\partial_{004}^*}{n}$$

$$E(e_0e_1)=\frac{1}{n}\partial_{220}^*, E(e_1e_2)=\frac{1}{n}\partial_{022}^*, E(e_0e_2)=\frac{1}{n}\partial_{202}^*$$

Isaki (1983) suggested ratio estimator t_1 for estimating S_y^2 as-

$$t_1=s_y^2\frac{S_y^2}{s_x^2}\text{; where } s_x^2 \text{ is unbiased estimator of } S_x^2 \tag{1.2}$$

Up to the first order of approximation, mean square error of t_1 is given by,

$$MSE(t_1)=\frac{S_y^4}{n}\left[\partial_{400}^*+\partial_{040}^*-2\partial_{220}^*\right] \tag{1.3}$$

Singh et al. (2007) proposed the exponential ratio-type estimator t_2 as-

$$t_2=s_y^2\exp\left[\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right] \tag{1.4}$$

And exponential product type estimator t_3 as-

$$t_3=s_y^2\left[\frac{s_x^2-S_x^2}{s_x^2+S_x^2}\right] \tag{1.5}$$

Following Kadilar and Cingi (2006), Singh et al. (2011) proposed an improved estimator for estimating population variance S_y^2, as-

$$t_4=s_y^2\left[k_4\exp\left\{\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right\}+(1-k_4)\left\{\frac{s_x^2-S_x^2}{s_x^2+S_x^2}\right\}\right] \tag{1.6}$$

where k_4 is a constant.

Up to the first order of approximation mean square errors of t_2, t_3 and t_4 are respectively given by

$$MSE(t_2)=\frac{S_y^4}{n}\left[\partial_{400}^*+\frac{\partial_{040}^*}{4}-\partial_{220}^*\right] \tag{1.7}$$

$$MSE(t_3) = \frac{S_y^4}{n}\left[\partial_{400}^* + \frac{\partial_{004}^*}{4} - \partial_{202}^*\right] \tag{1.8}$$

$$MSE(t_4) = \frac{S_y^4}{n}\left[\partial_{400}^* + k_4^2\frac{\partial_{040}^*}{4} + (1-k_4)^2\frac{\partial_{004}^*}{4} - k_4\partial_{220}^* + (1-k_4)\partial_{202}^* - \frac{k_4(1-k_4)}{2}\partial_{022}^*\right] \tag{1.9}$$

where $k_4 = \frac{(\partial_{004}^*/2) + \partial_{220}^* + \partial_{022}^*}{2(\partial_{040}^* + \partial_{004}^* + \partial_{022}^*}$.

3. Improved Estimator

Using Singh and Solanki (2011), we propose some improved estimators for estimating population variance S_y^2 as-

$$t_5 = s_y^2\left[\frac{cS_x^2 - Ds_x^2}{(c-d)S_x^2}\right]^p \tag{1.10}$$

$$t_6 = s_y^2\left[\frac{(a+b)S_z^2}{aS_x^2 + bs_x^2}\right]^q \tag{1.11}$$

$$t_7 = s_y^2\left[k_7\left\{\frac{cS_x^2 - Ds_x^2}{(c-d)S_x^2}\right\}^p + (1-k_7)\left\{\frac{(a+b)S_z^2}{aS_x^2 + bs_x^2}\right\}^q\right] \tag{1.12}$$

where a, b, c, d are suitably chosen constants and k_7 is a real constant to be determined so as to minimize MSE's.

Expressing t_5, t_6 and t_7 in terms of e_i's, we have

$$t_5 = S_y^2\left[1 - x_1pe_1 + e_0 - x_1pe_0e_1\right] \tag{1.13}$$

where, $x_1 = \frac{d}{(c-d)}$.

$$t_6 = S_y^2\left[1 - qx_2e_2 + e_0 - qx_2e_0e_2\right] \tag{1.14}$$

where, $x_2 = \frac{b}{(a+b)}$.

$$t_7 = S_y^2\left[1 + e_0 + px_1k_7(e_1' - e_1) + (k_7 - 1)qx_2e_2'\right] \tag{1.15}$$

The mean squared error of estimators are obtained by subtracting S_y^2 from each estimator and squaring both sides and than taking expectations-

$$MSE(t_5) = \frac{S_y^4}{n}\left[\partial^*_{400} + x_1^2 p^2 \partial^*_{040} - 2x_1 p \partial^*_{220}\right] \tag{1.16}$$

Differentiating (1.16) with respect to x_1, we get the optimum value of x_1 as-

$$x_{1(opt)} = \frac{\partial^*_{220}}{p\partial^*_{040}}.$$

$$MSE(t_6) = \frac{S_y^4}{n}\left[\partial^*_{400} + x_2^2 q^2 \partial^*_{004} - 2x_1 \partial^*_{202}\right] \tag{1.17}$$

Differentiating (1.17) with respect to x_2, we get the optimum value of x_2 as –

$$x_{2(opt)} = \frac{\partial^*_{202}}{q\partial^*_{004}}.$$

$$MSE(t_7) = \frac{S_y^4}{n}\left[A + k_7^2 B + (1-k_7)C - 2k_7(1-k_7)E - 2(1-k_7)F\right] \tag{1.18}$$

Differentiating (1.18) with respect to k_7, we get the optimum value k_7 of as –

$$k_{7(opt)} = \frac{C + D - F - E}{B + C - 2E}.$$

where

$A = \partial^*_{400}$, $B = x_1^2 p^2 \partial^*_{040}$,

$C = x_2^2 q^2 \partial^*_{004}$, $D = x_1 p \partial^*_{220}$,

$E = x_1 x_2 pq \partial^*_{022}$, $F = x_2 q \partial^*_{202}$.

2. Estimators In Two Phase Sampling

In certain practical situations when S_x^2 is not known a priori, the technique of two phase sampling or double sampling is used. Allowing SRSWOR design in each phase, the two – phase sampling scheme is as follows:

- The first phase sample $s'_n (s'_n \subset V)$ of a fixed size n' is drawn to measure only x and z in order to formulate the a good estimate of S_x^2 and S_z^2, respectively.

- Given s'_n ,the second phase sample $s_n (s_n \subset s'_n)$ of a fixed size n is drawn to measure y only.

Existing Estimators

Singh et al. (2007) proposed some estimators to estimate S_y^2 in two phase sampling, as:

$$t'_2 = s_y^2 \exp\left[\frac{s'^2_x - s_x^2}{s'^2_x + s_x^2}\right] \tag{2.1}$$

$$t'_3 = s_y^2 \exp\left[\frac{s'^2_z - s_z^2}{s'^2_z + s_z^2}\right] \tag{2.2}$$

$$t'_4 = s_y^2\left[k'_4 \exp\left\{\frac{s'^2_x - s_x^2}{s'^2_x + s_x^2}\right\} + (1-k'_4)\exp\left\{\frac{s'^2_z - s_z^2}{s'^2_z + s_z^2}\right\}\right] \tag{2.3}$$

MSE of the estimator t'_2, t'_3 and t'_4 are respectively, given by

$$MSE(t'_2) = S_y^4\left[\frac{\partial^*_{400}}{n} + \frac{1}{4}\left(\frac{1}{n} - \frac{1}{n'}\right)\partial^*_{040} + \left(\frac{1}{n'} - \frac{1}{n}\right)\partial^*_{220}\right] \tag{2.4}$$

$$MSE(t'_3) = S_y^4\left[\frac{\partial^*_{400}}{n} + \frac{1}{4}\left(\frac{1}{n} - \frac{1}{n'}\right)\partial^*_{004} - \left(\frac{1}{n'} - \frac{1}{n}\right)\partial^*_{202}\right] \tag{2.5}$$

$$MSE(t'_4) = S_y^4\left[A' + k'^2_4 B' + (1-k'_4)^2 C' + k'_4 D' + (1-k'_4)E'\right] \tag{2.6}$$

And $k'_{4(opt)} = \dfrac{2C' + E' - D'}{2(B' + C')}$

Where,

$$A' = \frac{\partial^*_{400}}{n}\ , B' = \frac{1}{4}\left(\frac{1}{n} - \frac{1}{n'}\right)\partial^*_{040}\ \ C' = \frac{1}{4n'}\partial^*_{004}$$

$$D' = \left(\frac{1}{n'} - \frac{1}{n}\right)\partial^*_{202}\ , E' = \frac{1}{n'}\partial^*_{202}$$

Proposed estimators in two phase sampling

The estimator proposed in section 3 will take the following form in two phase sampling;

$$t'_5 = s_y^2\left[\frac{cs'^2_x - ds_x^2}{(c-d)s'^2_x}\right] \tag{2.7}$$

$$t_6' = s_y^2\left[\frac{(a+b)S_z^2}{aS_z^2 + bs'^2_z}\right] \tag{2.8}$$

$$t_7' = s_y^2\left[k_7'\left\{\frac{cs'^2_x - ds_x^2}{(c-d)s'^2_x}\right\} + (1-k_7')\left\{\frac{(a+b)S_z^2}{aS_z^2 + bs'^2_z}\right\}\right] \tag{2.9}$$

Let,

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

$s_z^2 = S_z^2(1+e_2)$, $s'^2_z = S_z^2(1+e_2')$

Where,

$$s'^2_x = \frac{1}{(n'-1)}\sum_{i=1}^{n'}(x_i - \overline{x}')^2 \text{ , } s'^2_z = \frac{1}{(n'-1)}\sum_{i=1}^{n'}(z_i - \overline{z}')^2$$

and $\overline{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i , \overline{z}' = \frac{1}{n'}\sum_{i=1}^{n'} z_i$

Also,

$$E(e_1') = E(e_2') = 0$$

$$E(e_1'^2) = \frac{\partial^*_{040}}{n'}, E(e_2'^2) = \frac{\partial^*_{004}}{n'}$$

$$E(e_0 e_1') = \frac{1}{n'}\partial^*_{220} \text{ , } E(e_0 e_2') = \frac{1}{n'}\partial^*_{202} \text{ , } E(e_1 e_1') = \frac{1}{n'}\partial^*_{040}$$

$$E(e_2 e_2') = \frac{1}{n'}\partial^*_{004} \text{ , } E(e_1 e_2') = \frac{1}{n'}\partial^*_{022} \text{ , } E(e_1' e_2') = \frac{1}{n'}\partial^*_{022}$$

Writing estimators t_5', t_6' and t_7' in terms of e_i's we have ,respectively

$$t_5' = S_y^2\left[1 + e_0 + px_1(e_1' - e_1)\right] \tag{2.10}$$

$$t_6' = S_y^2\left[1 + e_0 - qx_2 e_2'\right] \tag{2.11}$$

$$t_7' = S_y^2\left[1 + px_1 k_7'((e_1' - e_1) + e_0 + (k_7' - 1)qx_2 e_2'\right] \tag{2.12}$$

Solving (1.10),(1.11) and (1.12),we get the MSE'S of the estimators t_5', t_6' and t_7', respectively as-

$$MSE(t_5') = S_y^4\left[\frac{\partial^*_{400}}{n} + p^2 x_1^2\left(\frac{1}{n} - \frac{1}{n'}\right)\partial^*_{040} + 2px_1\left(\frac{1}{n'} - \frac{1}{n}\right)\partial^*_{220}\right] \tag{2.13}$$

Differentiate (2.13) w.r.t. x_1, we get the optimum value of x_1 as-

$$x_{1(opt)} = \frac{\partial^*_{220}}{p\partial^*_{040}}$$

$$MSE(t'_6) = S_y^4\left[\frac{\partial^*_{400}}{n} + q^2 x_2^2 \frac{1}{n'}\partial^*_{004} + 2qx_2 \frac{1}{n'}\partial^*_{202}\right] \quad (2.14)$$

Differentiate (2.14) with respect to x_2, we get the optimum value of x_2 as –

$$x_{2(opt)} = \frac{\partial^*_{202}}{q\partial^*_{004}}.$$

$$MSE(t'_7) = S_y^4\left[A_1 + B_1 k_7'^2 + (k'_7 - 1)^2 C_1 + 2k'_7 D_1 + 2(k'_7 - 1)E_1\right] \quad (2.15)$$

Differentiate (2.15) with respect to k'_7, we get the optimum value of k'_7 as –

$$k'_{7(opt)} = \frac{C_1 - D_1 - E_1}{B_1 + C_1}$$

Where,

$$A_1 = \frac{\partial^*_{400}}{n},\; B_1 = p^2 x_1^2\left(\frac{1}{n} - \frac{1}{n'}\right)\partial^*_{040},\; C_1 = px_1\frac{q^2 x_2^2}{n'}\partial^*_{004}$$

$$D_1 = px_1\left(\frac{1}{n'} - \frac{1}{n}\right)\partial^*_{220},\; E_1 = qx_2\frac{\partial^*_{202}}{n'}$$

5. Empirical Study

In support of theoretical result an empirical study is carried out. The data is taken from Murthy(1967):

$$\partial_{400} = 3.726, \partial_{040} = 2.912, \partial_{044} = 2.808$$

$$\partial_{022} = 2.73, \partial_{202} = 2.979, \partial_{220} = 3.105$$

$$c_x = 0.5938, c_y = 0.7531, c_z = 0.7205$$

$$\rho_{yz} = 0.904, \rho_{xy} = 0.98, n = 7, n' = 15$$

$$\overline{X} = 747.5882, \overline{Y} = 199.4412, \overline{Z} = 208.8824$$

- In Table 5.1 percent relative efficiency of various estimators of S_y^2 is written with respect to s_y^2

- **Table 5.1: PRE of the estimator with respect to s_y^2**

Estimators	**PRE**
s_y^2	100
t_1	636.9158
t_2	248.0436
t_3	52.86019
t_4	**699.2526**
t_5	667.2895
t_6	486.9362
t_7	**699.5512**

- In Table 5.2 percent relative efficiency of various estimators of S_y^2 is written with respect to s_y^2 in two phase sampling:

- **Table 5.2: PRE of the estimators in two phase sampling with respect to s_y^2**

Estimators	**PRE**
s_y^2	100
t_2'	142.60
t_3'	66.42
t_4'	**460.75**
t_5'	182.95
t_6'	158.93
t_7'	**568.75**

6. Conclusion

In Table 5.1 and 5.2 percent relative efficiencies of various estimators are written with respect to s_y^2. From Table 5.1 and 5.2 we observe that the proposed estimator under optimum condition performs better than usual estimator, Isaki (1983) estimator and Singh et al. (2007) estimator.

7. References

Bahl, S. and Tuteja, R.K. (1991): Ratio and Product type exponential estimator. Information and Optimization sciences, Vol. XII, I, 159-163.

Grover, L. K . (2010): A correction note on improvement in variance estimation using auxiliary information. Commun. Stat. Theo. Meth. 39:753–764.

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

Kadilar,C. and Cingi,H. (2006) *:* Improvement in estimating the population mean in simple random sampling. Applied Mathematics Letters 19 (2006), 75–79.

Murthy, M. N. (1967).: Sampling Theory and Methods, Statistical Publishing Soc., Calcutta, India.

Singh, H. P. and Singh, R. (2001): Improved Ratio-type Estimator for Variance Using Auxiliary Information. J.I.S.A.S.,54(3),276-287.

Singh, H. P. and Singh, R. (2002): A class of chain ratio-type estimators for the coefficient of variation of finite population in two phase sampling. Aligarh Journal of Statistics, 22,1-9.

Singh, H. P. and Singh, R. (2003): Estimation of variance through regression approach in two phase sampling. Aligarh Journal of Statistics, 23, 13-30.

Singh, H. P, and Solanki, R. S. (2012): A new procedure for variance estimation in simple random sampling using auxiliary information. Statistical papers, DOI 10.1007/s00362-012-0445-2.

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2011): Improved exponential estimator for population variance using two auxiliary variables. Italian Jour. Of Pure and Applied. Math., 28, 103-110.

Singh, R., Kumar, M., Singh, A.K. and Smarandache, F. (2011): A family of estimators of population variance using information on auxiliary attribute. Studies in sampling techniques and time series analysis. Zip publishing, USA.

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2008): Almost unbiased ratio and product type estimator of finite population variance using the knowledge of kurtosis of an auxiliary variable in sample surveys. Octogon Mathematical Journal, Vol. 16, No. 1, 123-130.