Rajesh Singh, Pankaj Chauhan, Nirmala Sawan

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

Florentin Smarandache

University of New Mexico, USA

A General Family of Estimators for Estimating Population Variance Using Known Value of Some Population Parameter(s)

Published in:
Rajesh Singh, Pankaj Chauhan, Nirmala Sawan, Florentin Smarandache (Editors)
AUXILIARY INFORMATION AND A PRIORI VALUES IN CONSTRUCTION OF IMPROVED ESTIMATORS
Renaissance High Press, Ann Arbor, USA, 2007
(ISBN-10): 1-59973-046-4
(ISBN-13): 978-1-59973-046-2
pp. 65 - 73

Abstract

A general family of estimators for estimating the population variance of the variable under study, which make use of known value of certain population parameter(s), is proposed. Some well known estimators have been shown as particular member of this family. It has been shown that the suggested estimator is better than the usual unbiased estimator, Isaki's (1983) ratio estimator, Upadhyaya and Singh's (1999) estimator and Kadilar and Cingi (2006). An empirical study is carried out to illustrate the performance of the constructed estimator over others.

Keywords: Auxiliary information, variance estimator, bias, mean squared error.

1. Introduction

In manufacturing industries and pharmaceutical laboratories sometimes researchers are interested in the variation of their produce or yields (Ahmed et.al. (2003)). 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 drawn. Further let y and x denote the study and the auxiliary variables respectively.

Let $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{n}\left(y_i - \overline{Y}\right)^2$ denotes respectively the unknown population mean and population variance of the study character y. Assume that population size N is very large so that the finite population correction term is ignored. It is established fact that in most of the survey situations, auxiliary information is available (or may be made to be available diverting some of the resources) in one form or the other. If used intelligibly, this information may yield estimators better than those in which no auxiliary information is used.

Assume that a simple random sample of size n is drawn without replacement. The usual unbiased estimator of S_y^2 is

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \tag{1.1}$$

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

When the population mean square $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$ is known, Isaki (1983) proposed a ratio estimator for S_y^2 as

$$t_1 = \left(\frac{s_y^2}{s_x^2}\right) S_x^2 \tag{1.2}$$

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

Several authors have used prior value of certain population parameter(s) to find more precise estimates. The use of prior value of coefficient of kurtosis in estimating the population variance of study character y was first made by Singh et. al. (1973). Kadilar and Cingi (2006) proposed modified ratio estimators for the population variance using different combinations of known values of coefficient of skewness and coefficient of variation.

In this paper, under SRSWOR, we have suggested a general family of estimators for estimating the population variance S_y^2. The expressions of bias and mean-squared error (MSE), up to the first order of approximation, have been obtained. Some well known estimators have been shown as particular member of this family.

2. The suggested family of estimators

Motivated by Khoshnevisan et. al. (2007), we propose following ratio-type estimators for the population variance as

$$t = s_y^2 \frac{\left(aS_x^2 - b\right)}{[\alpha(as_x^2 - b) + (1-\alpha)(aS_x^2 - b)]} \tag{2.1}$$

where $(a \neq 0)$, b are either real numbers or the function of the known parameters of the auxiliary variable x such as coefficient of variation C(x) and coefficient of kurtosis $(\beta_2(x))$.

The following scheme presents some of the important known estimators of the population variance, which can be obtained by suitable choice of constants α, a and b:

Table 2.1 : Some members of the proposed family of the estimators 't'

Estimator	Values of		
	α	**a**	**b**
$t_0 = s_y^2$	0	0	0
$t_1 = \frac{s_y^2}{s_x^2} S_x^2$ Isaki (1983) estimator	1	1	0
$t_2 = \frac{s_y^2}{s_x^2 - C_x}[S_x^2 - C_x]$ Kadilar and Cingi (2006) estimator	1	1	C_x
$t_3 = \frac{s_y^2}{s_x^2 - \beta_2(x)}[S_x^2 - \beta_2(x)]$	1	1	$\beta_2(x)$
$t_4 = \frac{s_y^2}{s_x^2 \beta_2(x) - C_x}[S_x^2 \beta_2(x) - C_x]$	1	$\beta_2(x)$	C_x
$t_5 = \frac{s_y^2}{s_x^2 C_x - \beta_2(x)}[S_x^2 C_x - \beta_2(x)]$	1	C_x	$\beta_2(x)$
$t_6 = \frac{s_y^2}{s_x^2 + \beta_2(x)}[S_x^2 + \beta_2(x)]$ Upadhyaya and Singh (1999)	1	1	$-\beta_2(x)$

The MSE of proposed estimator 't' can be found by using the firs degree approximation in the Taylor series method defined by

$$MSE(t) \cong d \sum d' \tag{2.2}$$

where

$$h = \left[\left.\frac{\partial h(a,b)}{\partial a}\right|_{S_y^2, S_x^2} \quad \left.\frac{\partial h(a,b)}{\partial b}\right|_{S_y^2, S_x^2} \right]$$

$$\Sigma = \begin{bmatrix} V(s_y^2) & Cov(s_y^2, s_x^2) \\ Cov(s_x^2, s_y^2) & V(s_x^2) \end{bmatrix}.$$

Here h(a,b) = h(s_y^2, s_x^2) = t. According to this definition, we obtain 'd' for the proposed estimator, t, as follows:

$$d = \left[1 \quad -\frac{\alpha a S_y^2}{a S_x^2 + b} \right]$$

MSE of the proposed estimator t using (2.2) is given by

$$MSE(t) \cong V(s_y^2) - 2\alpha\left(\frac{aS_y^2}{aS_x^2 - b}\right)Cov(s_y^2, s_x^2) + \left(\frac{\alpha a S_y^2}{aS_x^2 - b}\right)V(s_x^2) \tag{2.3}$$

where

$$\left.\begin{aligned} V(s_y^2) &= \lambda S_y^4[\beta_2(y) - 1] \\ V(s_x^2) &= \lambda S_y^4[\beta_2(x) - 1] \\ Cov(s_y^2, s_x^2) &= \lambda S_y^2 S_x^2 (h - 1) \end{aligned}\right\} \tag{2.4}$$

where $\lambda = \frac{1}{n}$, $\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}$, $\beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}$, $h = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$,

$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{Y})^r (x_i - \overline{X})^s$, (r, s) being non negative integers.

Using (2.4), MSE of t can be written as

$$MSE(t) \cong \lambda S_y^4 \left\{\beta_2(y) - 1 - 2\alpha\theta(h-1) + \alpha^2\theta^2\left(\beta_2(x) - 1\right)\right\} \tag{2.5}$$

where $\theta = \frac{aS_x^2}{aS_x^2 - b}$.

The MSE equation of estimators listed in Table2.1 can be written as-

$$\mathrm{MSE}(t_i) \cong \lambda S_y^4 \left\{\beta_2(y) - 1 - 2\alpha\theta_i(h-1) + \alpha^2\theta_i^2\left(\beta_2(x)-1\right)\right\}, \quad i = 2,3,4,5,6 \tag{2.6}$$

where

$$\theta_2 = \frac{S_x^2}{S_x^2 - C_x}, \qquad \theta_3 = \frac{S_x^2}{S_x^2 - \beta_2(x)},$$

$$\theta_4 = \frac{S_x^2\beta_2(x)}{S_x^2\beta_2(x) - C_x}, \quad \theta_5 = \frac{S_x^2 C_x}{S_x^2 C_x - \beta_2(x)}, \qquad \theta_6 = \frac{S_x^2}{S_x^2 + \beta_2(x)}.$$

Minimization of (2.5) with respect to α yields its optimum value as

$$\alpha = \frac{C}{\theta} = \alpha_{opt} \tag{2.7}$$

where $C = \dfrac{(h-1)}{\{\beta_2(x) - 1\}}$.

By substituting α_{opt} in place of α in (2.5) we get the resulting minimum variance of t as

$$\min.\mathrm{MSE}(t) = \lambda S_y^4\left[\beta_2(y) - 1 - \{\beta_2(x) - 1\}\right] \tag{2.8}$$

3. Efficiency comparisons

Up to the first order of approximation, variance (ignoring finite population correction) of $t_o = s_y^2$ and t_1 is given by –

$$\mathrm{Var}(s_y^2) = \lambda S_y^4\left[\beta_2(y) - 1\right] \tag{3.1}$$

$$\mathrm{MSE}(t_1) = \lambda S_y^4\left[\{\beta_2(y) - 1\} + \{\beta_2(x) - 1\}(1 - 2C)\right] \tag{3.2}$$

From (2.6), (2.8), (3.1), and (3.2), we have

$$\mathrm{Var}(s_y^2) - \min.\mathrm{MSE}(t) = \lambda S_y^4\{\beta_2(x) - 1\}C^2 > 0 \tag{3.3}$$

$$\mathrm{MSE}(t_i) - \min.\mathrm{MSE}(t) = \lambda S_y^4\{\beta_2(x) - 1\}(\theta_i - C^2) > 0, \; i = 1,2,3,4,5,6 \tag{3.4}$$

provided $C \neq \theta_i$.

Thus it follows from (3.3) and (3.4) that the suggested estimator under 'optimum' condition is (i) always better then s_y^2, (ii) better than Isaki's (1983) estimator t_1 except when C = 1 in which both are equally efficient, and (iii) Kadilar and Cingi (2006) estimators $t_i (i = 2,3,4,5)$ except when $C = \theta_i (i = 2,3,4,5)$ in which t and $t_i (i = 2,3,4,5)$ are equally efficient.

4. Empirical study

We use data in Kadilar and Cingi (2004) to compare efficiencies between the traditional and proposed estimators in the simple random sampling.

In Table 4.1, we observe the statistics about the population.

Table 4.1: Data statistics of the population for the simple random sampling

N = 106, n = 20, $\rho = 0.82$, $C_y = 4.18$, $C_x = 2.02$, $\overline{Y} = 15.37$, $\overline{X} = 243.76$,
$S_y = 64.25$, $S_x = 491.89$, $\beta_2(x) = 25.71$, $\beta_2(y) = 80.13$, $\lambda = 0.05$, $\theta = 33.30$.

The percent relative efficiencies of the estimators s_y^2, $t_i (i = 2,3,4,5,6)$ and $\min.MSE(t)$ with respect to s_y^2 have been computed and presented in Table 4.2 below.

Table 4.2: Relative efficiencies (%) of s_y^2, $t_i(i = 2,3,4,5,6)$ and $\min.MSE(t)$ with respect to s_y^2.

Estimator	PRE $(., s_y^2)$
$t_0 = s_y^2$	100
t_1	201.6564
t_2	201.6582
t_3	201.6782
t_4	201.6565
t_5	201.6672
t_6	201.6347
$\min.MSE(t)$	214.3942

5. Conclusion

From theoretical discussion in section 3 and results of the numerical example, we infer that the proposed estimator ‘t’ under optimum condition performs better than usual estimator s_y^2, Isaki’s (1983) estimator t_1, Kadilar and Cingi’s (2006) estimators (t_2, t_3, t_4, t_5) and Upadhyaya and Singh’s (1999) estimator t_6 .

References

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

Isaki, C.T. (1983): Variance estimation using auxiliary information. Jour. Amer. Stat. Assoc., 78, 117-123.

Kadilar, C. and Cingi, H. (2006): Ratio estimators for the population variance in simple and stratified random sampling. Applied Mathematics and Computation 173 (2006) 1047-1059.

Singh, J., Pandey, B.N. and Hirano, K. (1973): On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. Ann. Inst. Statist. Math., 25, 51-55.

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.