Rajesh Singh, Mukesh Kumar
Department of Statistics, B.H.U., Varanasi (U.P.)-India

Ashish K. Singh
College of Management Studies,
Raj Kumar Goel Institute of Technology

Florentin Smarandache
Department of Mathematics, University of New Mexico, Gallup, USA

A Family of Estimators of Population Variance Using Information on Auxiliary Attribute

Published in:
Rajesh Singh, F. Smarandache (Editors)
STUDIES IN SAMPLING TECHNIQUES AND TIME SERIES ANALYSIS
Zip Publishing, Columbus, USA, 2011
ISBN 978-1-59973-159-9
pp. 63 - 70

Abstract

This chapter proposes some estimators for the population variance of the variable under study, which make use of information regarding the population proportion possessing certain attribute. Under simple random sampling without replacement (SRSWOR) scheme, the mean squared error (MSE) up to the first order of approximation is derived. The results have been illustrated numerically by taking some empirical population considered in the literature.

Keywords: Auxiliary attribute, exponential ratio-type estimates, simple random sampling, mean square error, efficiency.

1. Introduction

It is well known that the auxiliary information in the theory of sampling is used to increase the efficiency of estimator of population parameters. Out of many ratio, regression and product methods of estimation are good examples in this context. There exist situations when information is available in the form of attribute which is highly correlated with y. Taking into consideration the point biserial correlation coefficient between auxiliary attribute and study variable, several authors including Naik and

Gupta (1996), Jhajj et. al. (2006), Shabbir and Gupta (2007), Singh et. al. (2007, 2008) and Abd-Elfattah et. al. (2010) defined ratio estimators of population mean when the prior information of population proportion of units, possessing the same attribute is available.

In many situations, the problem of estimating the population variance σ^2 of study variable y assumes importance. When the prior information on parameters of auxiliary variable(s) is available, Das and Tripathi (1978), Isaki (1983), Prasad and Singh (1990), Kadilar and Cingi (2006, 2007) and Singh et. al. (2007) have suggested various estimators of S_y^2.

In this chapter we have proposed family of estimators for the population variance S_y^2 when one of the variables is in the form of attribute. For main results we confine ourselves to sampling scheme SRSWOR ignoring the finite population correction.

2. The proposed estimators and their properties

Following Isaki (1983), we propose a ratio estimator

$$t_1 = s_y^2 \frac{S_\phi^2}{s_\phi^2} \tag{2.1}$$

Next we propose regression estimator for the population variance

$$t_2 = s_y^2 + b\left(S_\phi^2 - s_\phi^2\right) \tag{2.2}$$

And following Singh et. al. (2009), we propose another estimator

$$t_3 = s_y^2 \exp\left[\frac{S_\phi^2 - s_\phi^2}{S_\phi^2 - s_\phi^2}\right] \tag{2.3}$$

where s_y^2 and s_ϕ^2 are unbiased estimator of population variances S_y^2 and S_ϕ^2 respectively and b is a constant, which makes the MSE of the estimator minimum.

To obtain the bias and MSE, we write-

$$s_y^2 = S_y^2(1+e_0), \qquad s_\phi^2 = S_\phi^2(1+e_1)$$

Such that $E(e_0) = E(e_1) = 0$

and $E(e_0^2) = \dfrac{(\delta_{40}-1)}{n}$, $\quad E(e_1^2) = \dfrac{(\delta_{04}-1)}{n}$, $\quad E(e_0 e_1) = \dfrac{(\delta_{22}-1)}{n}$,

where $\delta_{pq} = \dfrac{\mu_{pq}}{\left(\mu_{20}^{p/2}\mu_{02}^{q/2}\right)}$, $\quad \mu_{pq} = \dfrac{\sum_{i=1}^{N}(y_i - \overline{Y})^p(\phi_i - P)^q}{(N-1)}$.

$\beta_{2(y)} = \dfrac{\mu_{40}}{\mu_{02}^2} = \delta_{40}$ and $\beta_{2(\phi)} = \dfrac{\mu_{04}}{\mu_{02}^2} = \delta_{04}$

Let $\beta_{2(y)}^* = \beta_{2(y)} - 1, \beta_{2(\phi)}^* = \beta_{2(x)} - 1$, and $\delta_{pq}^* = \delta_{pq} - 1$

P is the proportions of units in the population.

Now the estimator t_1 defined in (2.1) can be written as

$$\left(t_1 - S_y^2\right) = S_y^2\left(e_0 - e_1 + e_1^2 - e_0 e_1\right) \tag{2.4}$$

Similarly, the estimator t_2 can be written as

$$\left(t_2 - S_y^2\right) = S_y^2 e_0 - bS_\phi^2 e_1 \tag{2.5}$$

And the estimator t_3 can be written as

$$\left(t_3 - S_y^2\right) = S_y^2\left(e_0 - \frac{e_1}{2} - \frac{e_0 e_1}{2} + \frac{3e_1^2}{8}\right) \tag{2.6}$$

The MSE of t_1, t_3 and variance of t_2 are given, respectively, as

$$MSE(t_{p1}) = \frac{S_y^4}{n}\left[\beta^*_{2(y)} + \beta^*_{2(\phi)} - 2\delta^*_{22}\right] \qquad (2.7)$$

$$MSE(t_{p3}) = \frac{S_y^4}{n}\left[\beta^*_{2(y)} + \frac{\beta^*_{2(\phi)}}{4} - \delta^*_{22}\right] \qquad (2.8)$$

The variance of t_{p2} is given as

$$V(t_2) = \frac{1}{n}\left[S_y^4(\lambda_{40} - 1) + b^2 S_\phi^2(\lambda_{04} - 1) - 2bS_y^2 S_x^2(\lambda_{22} - 1)\right] \qquad (2.9)$$

On differentiating (2.9) with respect to b and equating to zero we obtain

$$b = \frac{S_y^2(\delta_{22} - 1)}{S_x^2(\delta_{04} - 1)} \qquad (2.10)$$

Substituting the optimum value of b in (2.9), we get the minimum variance of the estimator t_2, as

$$\min.V(t_2) = \frac{S_y^4}{n}\beta^*_{2(y)}\left[1 - \frac{\delta^{*2}_{22}}{\beta^*_{2(y)}\beta^*_{2(\phi)}}\right] = Var\left(\hat{S}^2\right)\left(1 - \rho^2_{(S_y^2, S_\phi^2)}\right) \qquad (2.11)$$

3. Adapted estimator

We adapt the Shabbir and Gupta (2007) and Grover (2010) estimator, to the case when one of the variables is in the form of attribute and propose the estimator t_4

$$t_4 = \left[k_1 s_y^2 + k_2\left(S_\phi^2 - s_\phi^2\right)\right]\exp\left(\frac{S_\phi^2 - s_\phi^2}{S_\phi^2 + s_\phi^2}\right) \qquad (3.1)$$

where k_1 and k_2 are suitably chosen constants.

Expressing equation (3.1) in terms of e's and retaining only terms up to second degree of e's, we have:

$$t_4 = \left[k_1 S_y^2(1 + e_0) - k_2 s_\phi^2 e_1\right]\left[1 - \frac{e_1}{2} + \frac{3}{8}e_1^2\right] \qquad (3.2)$$

Up to first order of approximation, the mean square error of t_4 is

$$MSE(t_4) = E\left(t_4 - S_y^2\right)^2$$

$$= S_y^4\left[(k_1-1)^2 + \lambda k_1^2\left(\beta_2^*(y) + \beta_2^*(\phi) - 2\delta_{22}\right) + \lambda k_1\left(\delta_{22}^* - \frac{3}{4}\beta_2^*(\phi)\right)\right.$$

$$\left. + S_\phi^4 k_2^2 \lambda \beta_2^*(\phi) + 2\lambda S_y^2 S_x^2\left[k_1 k_2\left(\beta_2^*(x) - \delta_{22}^*\right) - \frac{k_2}{2}\beta_2^*(x)\right]\right] \quad (3.3)$$

where, $\lambda = \frac{1}{n}$

On partially differentiating (3.3) with respect to $k_i\,(i = 1,2)$, we get optimum values of k_1 and k_2, respectively as

$$k_1^* = \frac{\beta_2^*(\phi)\left(2 - \frac{\lambda}{4}\beta_2^*(\phi)\right)}{2\left(\beta_2^*(\phi)(\lambda A + 1) - \lambda B^2\right)} \quad (3.4)$$

and

$$k_2^* = \frac{S_y^2\left[\beta_2^*(\phi)(\lambda A + 1) - \lambda B^2 - B\left(2 - \frac{\lambda}{4}\beta_2^*(\phi)\right)\right]}{2S_x^2\left(\beta_2^x(\phi)(\lambda A + 1) - \lambda B^2\right)} \quad (3.5)$$

where,

$A = \beta_2^*(y) + \beta_2^*(\phi) - 2\delta_{22}^*$ and $B = \beta_2^*(\phi) - \delta_{22}^*$.

On substituting these optimum values of k_1 and k_2 in (3.3), we get the minimum value of MSE of t_4 as

$$MSE(t_4) = \frac{MSE(t_2)}{1 + \frac{MSE(t_2)}{S_y^4}} - \frac{\lambda\beta_2^*(x)\left(MSE(t_2) + \frac{\lambda S_y^4 \beta_2^*(\phi)}{16}\right)}{4\left(1 + \frac{MSE(t_2)}{S_y^4}\right)} \quad (3.6)$$

4. Efficiency Comparison

First we have compared the efficiency of proposed estimator under optimum condition with the usual estimator as -

$$V\left(\hat{S}_y^2\right) - MSE\left(\hat{S}_p^2\right)_{opt} = \frac{\lambda S_y^4 \delta_{22}^{*2}}{\beta_{2(x)}^*} - \frac{MSE(t_2)}{1+\frac{MSE(t_2)}{S_y^4}} + \frac{\lambda\beta_2^*(x)\left(MSE(t_2)+\frac{\lambda S_y^4 \beta_2^*(\phi)}{16}\right)}{4\left(1+\frac{MSE(t_2)}{S_y^4}\right)} \geq 0 \text{ always.} \qquad (4.1)$$

Next we have compared the efficiency of proposed estimator under optimum condition with the ratio estimator as -

From (2.1) and (3.6) we have

$$MSE(t_2) - MSE\left(\hat{S}_p^2\right)_{opt} = \lambda S_y^4\left[\sqrt{\beta_{2(x)}} - \frac{\delta_{22}^*}{\sqrt{\beta_{2(x)}^*}}\right]^2 - \frac{MSE(t_2)}{1+\frac{MSE(t_2)}{S_y^4}} + \frac{\lambda\beta_2^*(x)\left(MSE(t_2)+\frac{\lambda S_y^4 \beta_2^*(\phi)}{16}\right)}{4\left(1+\frac{MSE(t_2)}{S_y^4}\right)} \geq 0 \text{ always.} \qquad (4.2)$$

Next we have compared the efficiency of proposed estimator under optimum condition with the exponential ratio estimator as -

From (2.3) and (3.6) we have

$$MSE(t_3) - MSE(\hat{S}_p^2)_{opt} = \lambda S_y^4 \left[\sqrt{\beta_{2(x)}} - \frac{\delta_{22}^*}{2\sqrt{\beta_{2(x)}^*}} \right]^2 - \frac{MSE(t_2)}{1 + \frac{MSE(t_2)}{S_y^4}}$$

$$+ \frac{\lambda \beta_2^*(x) \left(MSE(t_2) + \frac{\lambda S_y^4 \beta_2^*(\phi)}{16} \right)}{4\left(1 + \frac{MSE(t_2)}{S_y^4}\right)} \geq 0 \text{ always.} \qquad (4.3)$$

Finally we have compared the efficiency of proposed estimator under optimum condition with the Regression estimator as -

$$MSE(t_2) - MSE(t_4) = \frac{MSE(t_2)}{1 + \frac{MSE(t_2)}{S_y^4}} - \frac{\lambda \beta_2^*(x) \left(MSE(t_2) + \frac{\lambda S_y^4 \beta_2^*(\phi)}{16} \right)}{4\left(1 + \frac{MSE(t_2)}{S_y^4}\right)} > 0 \text{ always.} \qquad (4.4)$$

5. Empirical study

We have used the data given in Sukhatme and Sukhatme (1970), p. 256.
Where, Y=Number of villages in the circle, and
ϕ Represent a circle consisting more than five villages.

n	N	S_y^2	S_p^2	λ_{40}	λ_{04}	λ_{22}
23	89	4.074	0.110	3.811	6.162	3.996

The following table shows PRE of different estimator's w. r. t. to usual estimator.

Table 1: PRE of different estimators

Estimators	t_0	t_1	t_2	t_3	t_4
PRE	100	141.898	262.187	254.274	**296.016**

Conclusion

Superiority of the proposed estimator is established theoretically by the universally true conditions derived in Sections 4. Results in Table 1 confirms this superiority numerically using the previously used data set.

References

Abd-Elfattah, A.M. El-Sherpieny, E.A. Mohamed, S.M. Abdou, O. F. (2010): Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Appl. Mathe. and Compt.

Das, A. K., Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance. Sankhya 40:139–148.

Grover, L.K. (2010): A Correction Note on Improvement in Variance Estimation Using Auxiliary Information. Communications in Statistics—Theory and Methods, 39: 753–764, 2010

Kadilar, C., Cingi, H. (2006). Improvement in variance estimation using auxiliary information. Hacettepe J. Math. Statist. 35(1):111–115.

Kadilar, C., Cingi, H. (2007). Improvement in variance estimation in simple random sampling. Commun. Statist. Theor. Meth. 36:2075–2081.

Isaki, C. T. (1983). Variance estimation using auxiliary information, Jour. of Amer. Statist.Asso.78, 117–123, 1983.

Jhajj, H.S., Sharma, M.K. and Grover, L.K. (2006). A family of estimators of Population mean using information on auxiliary attribute. Pak. J. Statist., 22(1),43-50.

Naik, V.D., Gupta, P.C. (1996): A note on estimation of mean with known population of an auxiliary character, Journal of Ind. Soci. Agri. Statist. 48(2) 151–158.

Prasad, B., Singh, H. P. (1990). Some improved ratio-type estimators of finite population variance in sample surveys. Commun. Statist. Theor. Meth. 19:1127–1139

Singh, R. Chauhan, P. Sawan, N. Smarandache, F. (2007): A general family of estimators for estimating population variance using known value of some population parameter(s). Renaissance High Press.

Singh, R. Chauhan, P. Sawan, N. Smarandache, F. (2008): Ratio estimators in simple random sampling using information on auxiliary attribute. Pak. J. Stat. Oper. Res. 4(1) 47–53

Shabbir, J., Gupta, S. (2007): On estimating the finite population mean with known population proportion of an auxiliary variable. Pak. Jour. of Statist.
23 (1) 1–9.

Shabbir, J., Gupta, S. (2007). On improvement in variance estimation using auxiliary information. Commun. Statist. Theor. Meth. 36(12):2177–2185.