Rajesh Singh, Mukesh Kumar, Sachin Malik, Manoj K. Chaudhary
Department of Statistics, B.H.U., Varanasi (U.P.)-India

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

A Family of Estimators for Estimating The Population Mean in Stratified Sampling

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. 55 - 62

Abstract

In this chapter, we have suggested an improved estimator for estimating the population mean in stratified sampling in presence of auxiliary information. The mean square error (MSE) of the proposed estimator have been derived under large sample approximation. Besides, considering the minimum case of the MSE equation, the efficient conditions between the proposed and existing estimators are obtained. These theoretical findings are supported by a numerical example.

Keywords : Auxiliary variable, mean square errors; exponential ratio type Estimates; stratified random sampling.

1. Introduction

In planning surveys, stratified sampling has often proved useful in improving the precision of other unstratified sampling strategies to estimate the finite population mean

$$\overline{Y} = \left(\sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi}\right)/N$$

Consider a finite population of size N. Let y and x respectively, be the study and auxiliary variates on each unit U_j (j=1,2,3…N) of the population U. Let the population be divided in to L strata with the h^{th}stratum containing N_h units, h=1,2,3…,L so that

$\sum_{h=1}^{L} N_h = N$. Suppose that a simple random sample of size n_h is drawn without replacement (SRSWOR) from the h^{th} stratum such that $\sum_{h=1}^{L} n_h = n$.

When the population mean $\bar{X}$ of the auxiliary variable x is known, Hansen et. al. (1946) suggested a "combined ratio estimator "

$$\bar{y}_{CR} = \bar{y}_{st}\left(\frac{\bar{X}}{\bar{x}_{st}}\right) \tag{1.1}$$

where, $\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h$, $\bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h$

$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ and $\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$

$w_h = \frac{N_h}{N}$ and $\bar{X} = \sum_{h=1}^{L} w_h \bar{X}_h$.

The "combined product estimator " for $\bar{Y}$ is defined by

$$\bar{y}_{CP} = \bar{y}_{st}\left(\frac{\bar{x}_{st}}{\bar{X}}\right) \tag{1.2}$$

To the first degree of approximation, the mean square error (MSE) of $\bar{y}_{CR}$ and $\bar{y}_{CP}$ are respectively given by –

$$MSE(\bar{y}_{CR}) \cong \sum_{i=1}^{L} w_h^2 \theta_h \left[S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{yxh}\right] \tag{1.3}$$

$$MSE(\bar{y}_{CP}) \cong \sum_{i=1}^{L} w_h^2 \theta_h \left[S_{yh}^2 + R^2 S_{xh}^2 + 2RS_{yxh}\right] \tag{1.4}$$

where $\theta_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$, $R = \frac{\bar{Y}}{\bar{X}}$ is the population ratio, S_{yh}^2 is the population variance of variate of interest in stratum h, S_{xh}^2 is the population variance of auxiliary variate in stratum h and S_{yxh} is the population covariance between auxiliary variate and variate of interest in stratum h.

Following Bahl and Tuteja (1991), Singh et. al. (2009) proposed following estimator in stratified random sampling -

$$\bar{y}_{er} = \bar{y}_{st} \exp\left[\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}}\right] \tag{1.5}$$

The MSE of $\bar{y}_{er}$, to the first degree of approximation is given by

$$MSE(\bar{y}_{er}) \cong \sum_{i=1}^{L} w_h^2 \theta_h \left[S_{yh}^2 + \frac{R^2}{4} S_{xh}^2 - RS_{yxh}\right] \tag{1.6}$$

Using the estimator $\bar{y}_{CR}$ and $\bar{y}_{CP}$, Singh and Vishwakarma (2005) suggested the combined ratio-product estimator for estimating $\bar{Y}$ as

$$\bar{y}_{RPC} = \bar{y}_{st}\left[\alpha \frac{\bar{X}}{\bar{x}_{st}} + (1-\alpha)\frac{\bar{x}_{st}}{\bar{X}}\right] \tag{1.7}$$

For minimum value of $\alpha = \frac{1}{2}(1 + C^*) = \alpha_0$ (say), the minimum MSE of the estimator $\bar{y}_{RPC}$ is given by

$$MSE\ (\bar{y}_{RPC}) = \sum_{i=1}^{L} w_h^2 \theta_h (1 - \rho^{*2})\ S_{yh}^2 \tag{1.8}$$

where $C^* = \frac{cov(\bar{y}_{st}, \bar{x}_{st})}{RV(\bar{x}_{st})}$, $\rho^* = \frac{cov(\bar{y}_{st}, \bar{x}_{st})}{R\sqrt{V(\bar{y}_{st})V(\bar{x}_{st})}}$, $R = \frac{\bar{Y}}{\bar{X}}$.

2. Proposed estimator

Following Singh and Vishwakarma (2005), we propose a new family of estimators -

$$t = \lambda\left[\bar{y}_{st}\exp\left[\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right]^{\alpha}\left(\frac{\bar{X}}{\bar{x}_{st}}\right)^{\beta}\right] + (1-\lambda)\left[\bar{y}_{st}\exp\left[\frac{\bar{x}_{st}-\bar{X}}{\bar{x}_{st}+\bar{x}}\right]^{\alpha}\left(\frac{\bar{x}_{st}}{\bar{X}}\right)^{\beta}\right] \tag{2.1}$$

where λ is real constant to be determined such that the MSE of t is a minimum and α, β are real constants such that $\beta = 1 - \alpha$.

Remark 2.1: For $\lambda = 1$ and $\alpha = 1$ the estimator t tends to Singh et. al. (2009) estimator. For $\lambda = 1$ and $\alpha = 0$ the estimator t takes the form of Hansen et. al. (1946) estimator $\bar{y}_{CR}$. For $\lambda = 0$ and $\alpha = 1$ the estimator t tends to Singh et. al. (2009) estimator. For $\lambda = 1$ and $\alpha = 0$ the estimator t takes the form of the estimator $\bar{y}_{CP}$.

To obtain the MSE of t to the first degree of approximation, we write

$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h = \bar{Y}(1 + e_0)$ and

$\bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h = \bar{X}(1 + e_1)$

Such that,

$E(e_0) = E(e_0) = 0.$

Under SRSWOR, we have

$E(e_0^2) = \frac{1}{\bar{Y}^2} \sum_{i=1}^{L} w_h^2 \theta_h S_{yh}^2$

$E(e_1^2) = \frac{1}{\bar{X}^2} \sum_{i=1}^{L} w_h^2 \theta_h S_{xh}^2$

$E(e_0 e_1) = \frac{1}{\bar{Y}\bar{X}} \sum_{i=1}^{L} w_h^2 \theta_h S_{yxh}$

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

$$t = \bar{Y}(1 + e_0)\left[\lambda\left\{\exp\left(\frac{-e_1}{2}\left(1 + \frac{e_1}{2}\right)^{-1}\right)\right\}^{\alpha}\{(1 + e_1)^{-1}\}^{1-\alpha} + (1 - \lambda)\left\{\exp\left(\frac{e_1}{2}\left(1 + \frac{e_1}{2}\right)^{-1}\right)\right\}^{1-\alpha}(1 + e_1)^{(1-\alpha)}\right] \tag{2.2}$$

We now assume that $|e_1| < 1$ so that we may expand $(1 + e_1)^{-1}$ as a series in powers of e_1. Expanding the right hand side of (2.2) to the first order of approximation, we obtain

$$(t - \bar{Y}) \cong \bar{Y}\left[e_0 + e_1\left(1 + \alpha\lambda - \frac{\alpha}{2} - 2\lambda\right)\right] \tag{2.3}$$

Squaring both sides of (2.3) and then taking expectations, we get the MSE of the estimator t, to the first order of approximation, as

$$MSE(t) = V(\bar{y}_{st}) + R^2(1 - 2\lambda)S_{xh}^2\{(1 - 2\lambda)A^2 + 2C^*A\} \tag{2.4}$$

where $A = \left(1 - \frac{\alpha}{2}\right)$.

Minimisation of (2.4) with respect to λ yields its optimum values as

$$\lambda_{opt} = \frac{1}{2}\left(1 + \frac{C^*}{A}\right) = \lambda_0(\text{say}) \tag{2.5}$$

Putting $\lambda = \lambda_0$ in (2.4) we get the minimum MSE of the estimator t as –

$$\min \text{MSE}(t) = V(\bar{y}_{st})(1-\rho^{*2})$$
$$= \sum_{i=1}^{L} w_h^2 \theta_h (1-\rho^{*2}) S_{yh}^2. \tag{2.6}$$

3. Efficiency comparisons

In this section we have compared proposed estimator with different already proposed estimators, obtained the conditions under which our proposed estimator performs better than other estimators.

First we have compared proposed estimator with simple mean in stratified random sampling.

$\text{MSE}(t) \leq \text{MSE}(\bar{y}_{st})$, if

$V(\bar{y}_{st}) + R^2(1-2\lambda)S_{xh}^2\{(1-2\lambda)A^2 + 2C^*A\} \leq V(\bar{y}_{st})$

$\min\left(\frac{1}{2} - \frac{1}{2} + \frac{C^*}{4}\right) \leq \lambda \leq \max\left(\frac{1}{2} - \frac{1}{2} + \frac{C^*}{4}\right)$

Next we compare proposed estimator with combined ratio estimator –

$\text{MSE}(t) \leq \text{MSE}(\bar{y}_{CR})$, if

$$V(\bar{y}_{st}) + \sum_{i=1}^{L} w_h^2 \theta_h R^2(1-2\lambda)S_{xh}^2\{(1-2\lambda)A^2 + 2C^*A\} \leq \sum_{i=1}^{L} w_h^2 \theta_h \left[S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{yxh}\right]$$

or, if $(1-2C^*) - (1-2\lambda)\big((1-2\lambda)A^2 + 2C^*A\big) \geq 0$

or, if $\frac{1}{2}\left\{\frac{A+1}{A}\right\} \leq \lambda \leq \frac{1}{2}\left\{\frac{2C^*+A-1}{A}\right\}$.

Next we compare efficiency of proposed estimator with product estimator

$\text{MSE}(t) \leq \text{MSE}(\bar{y}_{PR})$, if

$$V(\bar{y}_{st}) + \sum_{i=1}^{L} w_h^2 \theta_h R^2(1-2\lambda)S_{xh}^2\{(1-2\lambda)A^2 + 2C^*A\} \leq \sum_{i=1}^{L} w_h^2 \theta_h \left[S_{yh}^2 + R^2 S_{xh}^2 + 2RS_{yxh}\right]$$

or, if $(1 + 2C^*) - (1 - 2\lambda)((1 - 2\lambda)A^2 + 2C^*A) \geq 0$

or, if $\frac{1}{2}\left\{\frac{A-1}{A}\right\} \leq \lambda \leq \frac{1}{2}\left\{\frac{2C^*+A+1}{A}\right\}$.

Next we compare efficiency of proposed estimator and exponential ratio estimator in stratified sampling

$MSE(t) \leq MSE(\bar{y}_{ER})$, if

$$V(\bar{y}_{st}) + \sum_{i=1}^{L} w_h^2\theta_h R^2(1 - 2\lambda)S_{xh}^2\{(1 - 2\lambda)A^2 + 2C^*A\} \leq \sum_{i=1}^{L} w_h^2\theta_h \left[S_{yh}^2 + \frac{R^2}{4}S_{xh}^2 - RS_{yxh}\right]$$

or, if $(1 - 4C^*) - 4(1 - 2\lambda)((1 - 2\lambda)A^2 + 2C^*A) \geq 0$

or, if $\frac{1}{2}\left\{\frac{1-2A}{2A}\right\} \leq \lambda \leq \frac{1}{2}\left\{\frac{4C^*+2A-1}{2A}\right\}$

Finally we compare efficiency of proposed estimator with exponential product estimator in stratified random sampling

$MSE(t) \leq MSE(\bar{y}_{EP})$, if

or, if $$V(\bar{y}_{st}) + \sum_{i=1}^{L} w_h^2\theta_h R^2(1 - 2\lambda)S_{xh}^2\{(1 - 2\lambda)A^2 + 2C^*A\} \leq \sum_{i=1}^{L} w_h^2\theta_h \left[S_{yh}^2 + \frac{R^2}{4}S_{xh}^2 + RS_{yxh}\right]$$

or, if $(1 + 4C^*) - 4(1 - 2\lambda)((1 - 2\lambda)A^2 + 2C^*A) \geq 0$

or, if $\frac{1}{2}\left\{\frac{-1-2A}{2A}\right\} \leq \lambda \leq \frac{1}{2}\left\{\frac{4C^*+2A+1}{2A}\right\}$

Whenever above conditions are satisfied the proposed estimator performs better than other mentioned estimators.

4. Numerical illustration

All the theoretical results are supported by using the data given in Singh and Vishwakarma (2005).

Data statistics:

Stratum	w_h	θ_h	S^2_{xh}	S^2_{yh}	S_{yxh}
2	0.5227	0.12454	132.66	259113.71	5709.16
3	0.2428	0.08902	38.44	65885.60	1404.71

R=49.03 and $\lambda_{opt} = 0.9422(\alpha = 0)$ and $1.384525\ (\alpha = 1)$

Using the above data percentage relative efficiencies of different estimators $\bar{y}_{CR}$, $\bar{y}_{CP}$, $\bar{y}_{ER}$, $\bar{y}_{EP}$ and proposed estimator t w.r.t $\bar{y}_{st}$ have been calculated.

Table 4.1: PRE of different estimators of $\bar{Y}$

Estimator	$\bar{y}_{st}$	$\bar{y}_{CR}$	$\bar{y}_{CP}$	$\bar{y}_{ER}$	$\bar{y}_{EP}$	$\bar{y}_{HPS(opt)}$	$\bar{y}_{PRP(opt)}$
PRE	100	1148.256	23.326	405.222	42.612	1403.317	1403.317

We have also shown the range of λ for which proposed estimator performs better than $\bar{y}_{st}$.

Table 4.2: Range of λ for which proposed estimator performs better than $\bar{y}_{st}$

Value of constant α	Form of proposed estimator	Range of λ
$\alpha = 0$	$\bar{y}_{HPS}$	(0.5,1.3)
$\alpha = 1$	$\bar{y}_{CER}$	(0.5,2.2)

5. Conclusion

From the theoretical discussion and empirical study we conclude that the proposed estimator under optimum conditions performs better than other estimators considered in the article. The relative efficiency of various estimators are listed in Table

4.1 and the range of λ for which proposed estimator performs better than $\overline{y}_{st}$ is written in Table 4.2.

References

Bahl. S. and Tuteja, R.K. (1991): Ratio and Product Type Exponential Estimator. Infrm. and Optim. Sci., XIII, 159-163.

Hansen, M.H. and Hurwitz, W.N. (1946): The problem of non-response in sample surveys. J. Am. Stat. Assoc. 41:517–529.

Singh, H. P. and Vishwakarma, G. K. (2005): Combined Ratio-Product Estimator of Finite Population Mean in Stratified Sampling. *Metodologia de Encuestas* 8: 35-44.

Singh, R., Kumar, M., Chaudhary, M.K., Kadilar, C. (2009) : Improved Exponential Estimator in Stratified Random Sampling. Pak. J. Stat. Oper. Res. 5(2), pp 67-82.