

5-1-2011

# Improved Estimation of the Population Mean Using Known Parameters of an Auxiliary Variable

Rajesh Tailor

*Vikram University, Ujjain, M.P., India, tailorraj@gmail.com*

Balkishan Sharma

*Vikram University, Ujjain, M.P., India*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Tailor, Rajesh and Sharma, Balkishan (2011) "Improved Estimation of the Population Mean Using Known Parameters of an Auxiliary Variable," *Journal of Modern Applied Statistical Methods*: Vol. 10 : Iss. 1 , Article 7.

DOI: 10.22237/jmasm/1304222760

## Improved Estimation of the Population Mean Using Known Parameters of an Auxiliary Variable

Rajesh Tailor    Balkishan Sharma  
Vikram University  
Ujjain, M.P., India

---

An improved ratio-cum-product type estimator of the finite population mean is proposed using known information on the coefficient of variation of an auxiliary variate and correlation coefficient between a study variate and an auxiliary variate. Realistic conditions are obtained under which the proposed estimator is more efficient than the simple mean estimator, usual ratio and product estimators and estimators proposed by Singh and Diwivedi (1981), Pandey and Dubey (1988), Upadhaya and Singh (1999), and Singh, et al., (2004). An empirical study supports theoretical findings.

Key words: Study variate, auxiliary variate, population mean, correlation coefficient, coefficient of variation.

---

### Introduction

Auxiliary information is frequently used at the estimation stage in order to improve the efficiency of the estimator(s) of the parameter(s) of a variate under study; ratio, product and regression methods of estimation are examples. When the correlation between study variate and the auxiliary variate is positive (high), the ratio method of estimation is used for estimating the population mean. Conversely, if the correlation is negative, the product method of estimation is preferred.

Consider a finite population  $U = (U_1, U_2, \dots, U_N)$  of  $N$  units. Let  $y_i$  and  $x_i$  be the values of the study variate  $y$  and auxiliary variate  $x$  respectively on the  $i^{\text{th}}$  unit  $U_i$  ( $i=1,2,3,\dots,N$ ). For estimating the population mean,  $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$ , of the study variate  $y$ , a

simple random sample of size  $n$  is drawn using the simple random sampling without replacement (SRSWOR) technique from  $U$ .

When the population mean  $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$ , of the auxiliary variate  $x$  is known, the classical ratio and product estimators of  $\bar{Y}$  are respectively defined by the ratio estimator

$$\hat{Y}_1 = \bar{y} \left( \frac{\bar{X}}{\bar{x}} \right) \quad (1.1)$$

and the product estimator

$$\hat{Y}_2 = \bar{y} \left( \frac{\bar{x}}{\bar{X}} \right) \quad (1.2)$$

where  $\bar{y} = \sum_{i=1}^n y_i / n$  and  $\bar{x} = \sum_{i=1}^n x_i / n$  are the sample means of  $y$  and  $x$  respectively based on  $n$  observations.

When the population mean  $\bar{X}$  and coefficient of variation ( $C_x$ ) of auxiliary variate  $x$  are known, Sisodia and Dwivedi (1981) suggested using a ratio type estimator for  $\bar{Y}$  as

---

Rajesh Tailor is a Reader in the School of Studies in Statistics. His research interests are in the field of sampling techniques. Email: tailorraj@gmail.com. Balkishan Sharma is a research scholar working on his Ph.D. under the supervision of Dr. Rajesh Tailor.

## IMPROVED ESTIMATION OF POPULATION MEAN

$$\hat{Y}_3 = \bar{y} \left( \frac{\bar{X} + C_x}{\bar{x} + C_x} \right). \quad (1.3)$$

Using the same of information, Pandey and Dubey (1988) suggested a product type estimator for  $\bar{Y}$  as

$$\hat{Y}_4 = \bar{y} \left( \frac{\bar{x} + C_x}{\bar{X} + C_x} \right) \quad (1.4)$$

Further, when the population mean  $\bar{X}$  of x and the correlation coefficient ( $\rho$ ) between y and x are known, Singh and Tailor (2003) suggested ratio and product type estimators for  $\bar{Y}$  respectively as

$$\hat{Y}_5 = \bar{y} \left( \frac{\bar{X} + \rho}{\bar{x} + \rho} \right) \quad (1.5)$$

and

$$\hat{Y}_6 = \bar{y} \left( \frac{\bar{x} + \rho}{\bar{X} + \rho} \right) \quad (1.6)$$

Kadilar and Singi (2006) suggested a ratio-type and a product type estimator for  $\bar{Y}$ , using coefficient of variation  $C_x$  and correlation coefficient ( $\rho$ ), as

$$\hat{Y}_7 = \bar{y} \left( \frac{\bar{X}C_x + \rho}{\bar{x}C_x + \rho} \right) \quad (1.7)$$

This study proposes a ratio-cum-product estimator utilizing the knowledge on  $\bar{X}$ ,  $C_x$  and  $\rho$  and its properties are examined.

### Proposed Ratio-Cum-Product Estimator

Motivated by Singh and Tailor (2005), the proposed ratio-cum-product estimator for  $\bar{Y}$  is

$$\hat{Y}_B = \bar{y} \left[ \alpha \left( \frac{\bar{X}C_x + \rho}{\bar{x}C_x + \rho} \right) + (1 - \alpha) \left( \frac{\bar{x}C_x + \rho}{\bar{X}C_x + \rho} \right) \right] \quad (2.1)$$

where  $\alpha$  is a suitably chosen scalar. It should be noted that  $\alpha = 1$ ,  $\hat{Y}_B$  reduces to the estimator  $\hat{Y}_7$  suggested by Kadilar and Cingi (2006) and for the  $\alpha = 0$  product version of the  $\hat{Y}_7$ . Thus, these two estimators are particular cases of the proposed estimator  $\hat{Y}_B$ . To obtain the bias and MSE of  $\hat{Y}_B$ ,  $\bar{y} = \bar{Y}(1 + e_0)$  and  $\bar{x} = \bar{X}(1 + e_1)$  such that  $E(e_0) = E(e_1) = 0$  and

$$E(e_0^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_y^2,$$

$$E(e_1^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_x^2,$$

$$E(e_0 e_1) = \left( \frac{1}{n} - \frac{1}{N} \right) \rho C_y C_x$$

where

$$C_y = S_y / \bar{Y},$$

$$C_x = S_x / \bar{X},$$

$$\rho = S_{yx} / S_x S_y,$$

$$K = \rho C_y / C_x,$$

and

$$S_x^2 = \sum_{i=1}^N (x_i - \bar{X})^2 / (N - 1),$$

$$S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 / (N - 1),$$

$$S_{xy} = \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X}) / (N - 1).$$

Expressing (2.1) in terms of  $e_i$ 's results in

$$\hat{Y}_B = \bar{Y}(1 + e_0) \left[ \alpha(1 + \lambda_3 e_1)^{-1} + (1 - \alpha)(1 + \lambda_3 e_1) \right], \quad (2.2)$$

where

$$\lambda_3 = \bar{X}C_x / (\bar{X}C_x + \rho).$$

To the first degree of approximation, the bias and mean squared error of  $\hat{Y}_B$  respectively are

$$B(\hat{Y}_B) = \frac{(1-f)}{n} \bar{Y} \lambda_3 C_x^2 [K + \alpha(\lambda_3 - 2K)], \quad (2.3)$$

and

$$MSE(\hat{Y}_B) = \frac{(1-f)}{n} \bar{Y}^2 [C_y^2 + (1-2\alpha)\lambda_3 C_x^2 \{(1-2\alpha)\lambda_3 + 2K\}]. \quad (2.4)$$

Thus, with  $\alpha = K/(2K - \lambda_3)$ , the estimator  $\hat{Y}_B$  is almost unbiased. It is also observed from (2.3) that the bias of  $\hat{Y}_B$  is negligible for large sample.

The mean squared error of  $\hat{Y}_B$  in (2.4) is minimized for

$$\alpha = \frac{(\lambda_3 + K)}{2\lambda_3} = \alpha_0. \quad (2.5)$$

Substitution of (2.5) in (2.1) yields the asymptotically optimum estimator (AOE) for  $\bar{Y}$  as

$$\hat{Y}_B^{(opt)} = \frac{\bar{y}}{2\lambda_3} \left[ (\lambda_3 + K) \left( \frac{\bar{X}C_x + \rho}{\bar{x}C_x + \rho} \right) + (\lambda_3 - K) \left( \frac{\bar{x}C_x + \rho}{\bar{X}C_x + \rho} \right) \right] \quad (2.6)$$

and placing (2.5) in (2.3) and (2.4), results in the bias and variance of  $\hat{Y}_B^{(opt)}$  respectively as

$$B(\hat{Y}_B^{(opt)}) = \frac{(1-f)}{2n} \bar{Y} C_x^2 (\lambda_3 - K)(\lambda_3 + 2K), \quad (2.7)$$

and

$$MSE(\hat{Y}_B^{(opt)}) = \frac{(1-f)}{n} S_y^2 (1 - \rho^2). \quad (2.8)$$

It is clear that mean squared error of  $\hat{Y}_B^{(opt)}$  is the same as that of the approximate variance of the usual linear regression estimator  $\bar{y}_{lr} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x})$ , where  $\hat{\beta}$  is the sample regression coefficient of y on x.

#### Efficiency Comparisons

Under simple random sampling without replacement (SRSWOR), the variance of sample mean  $\bar{y}$  is

$$V(\bar{y}) = \frac{(1-f)}{n} \bar{Y}^2 C_y^2 \quad (3.1)$$

and the mean squared error of  $\hat{Y}_i$  (i=1 to 8) to the first degree of approximation are respectively given by:

$$MSE(\hat{Y}_1) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 - 2K)] \quad (3.2)$$

$$MSE(\hat{Y}_2) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 + 2K)] \quad (3.3)$$

$$MSE(\hat{Y}_3) = \theta \bar{Y}^2 [C_y^2 + \lambda_1 C_x^2 (\lambda_1 - 2K)] \quad (3.4)$$

$$MSE(\hat{Y}_4) = \theta \bar{Y}^2 [C_y^2 + \lambda_1 C_x^2 (\lambda_1 + 2K)] \quad (3.5)$$

$$MSE(\hat{Y}_5) = \theta \bar{Y}^2 [C_y^2 + \lambda_2 C_x^2 (\lambda_2 - 2K)] \quad (3.6)$$

$$MSE(\hat{Y}_6) = \theta \bar{Y}^2 [C_y^2 + \lambda_2 C_x^2 (\lambda_2 + 2K)] \quad (3.7)$$

$$MSE(\hat{Y}_7) = \theta \bar{Y}^2 [C_y^2 + \lambda_3 C_x^2 (\lambda_3 - 2K)] \quad (3.8)$$

where

$$\lambda_1 = \frac{\bar{X}}{\bar{X} + C_x},$$

$$\lambda_2 = \frac{\bar{X}}{\bar{X} + \rho},$$

IMPROVED ESTIMATION OF POPULATION MEAN

$$\lambda_3 = \frac{\overline{XC}_x}{\overline{XC}_x + \rho},$$

$$C_y = S_y / \bar{Y},$$

$$\theta = \left( \frac{1}{n} - \frac{1}{N} \right),$$

$$K = \frac{\rho C_y}{C_x}$$

and

$$S_y^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 / (N-1).$$

From (2.4) and (3.1), it is observed that  $\hat{Y}_B$  is more efficient than the usual unbiased estimator  $\bar{y}$  if:

$$\left. \begin{array}{l} \text{either } \frac{1}{2} < \alpha < \left( \frac{1+K}{2+\lambda_3} \right) \\ \text{or } \left( \frac{1+K}{2+\lambda_3} \right) < \alpha < \frac{1}{2} \end{array} \right\} \quad (3.9)$$

A comparison of (2.4) and (3.2) shows that  $\hat{Y}_B$  is more efficient than the usual ratio estimator  $\hat{Y}_1$  if:

$$\left. \begin{array}{l} \text{either } \frac{(1+\lambda_3)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3+2K-1}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3+2K-1}{2\lambda_3} \right) < \alpha < \frac{(1+\lambda_3)}{2\lambda_3} \end{array} \right\} \quad (3.10)$$

From (2.4) and (3.3) it is clear that  $\hat{Y}_B$  would be more efficient than  $\hat{Y}_2$  if:

$$\left. \begin{array}{l} \text{either } \left( \frac{\lambda_3+2K+1}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3-1)}{2\lambda_3} \\ \text{or } \frac{(\lambda_3-1)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3+2K+1}{2\lambda_3} \right) \end{array} \right\} \quad (3.11)$$

Comparing (2.4) and (3.4), it is observed that  $\hat{Y}_B$  is more efficient than the Sisodia and Dwivedi (1981) estimator  $\hat{Y}_3$  if

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3+\lambda_1)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3-\lambda_1+2K}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3-\lambda_1+2K}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3+\lambda_1)}{2\lambda_3} \end{array} \right\} \quad (3.12)$$

Comparing (2.4) and (3.5), it is observed that  $\hat{Y}_B$  is more efficient than the Pandey and Dubey (1988) estimator  $\hat{Y}_4$  if

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3-\lambda_1)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3+\lambda_1+2K}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3+\lambda_1+2K}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3-\lambda_1)}{2\lambda_3} \end{array} \right\} \quad (3.13)$$

Comparing (2.4) and (3.6), conditions under which suggested estimator  $\hat{Y}_B$  is more efficient than the Singh and Tailor (2003) ratio type estimator  $\hat{Y}_5$  when

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3+\lambda_2)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3-\lambda_2+2K}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3-\lambda_2+2K}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3+\lambda_2)}{2\lambda_3} \end{array} \right\} \quad (3.14)$$

Similarly conditions under which suggested estimator  $\hat{Y}_B$  is more efficient than the Singh and Tailor (2005) product type estimator  $\hat{Y}_6$  when

$$\left. \begin{array}{l} \text{either } \frac{(\lambda_3 - \lambda_2)}{2\lambda_3} < \alpha < \left( \frac{\lambda_3 + 2K + \lambda_2}{2\lambda_3} \right) \\ \text{or } \left( \frac{\lambda_3 + 2K + \lambda_2}{2\lambda_3} \right) < \alpha < \frac{(\lambda_3 - \lambda_2)}{2\lambda_3} \end{array} \right\} \quad (3.15)$$

Comparing (2.4) and (3.8), it is observed that  $\hat{Y}_B$  is more efficient than the Kadilar and Cingi (2006) ratio type estimator  $\hat{Y}_7$ , if

$$\left. \begin{array}{l} \text{either } \frac{1}{2} < \alpha < \frac{K}{\lambda_3} \\ \text{or } \frac{K}{\lambda_3} < \alpha < \frac{1}{2} \end{array} \right\} \quad (3.16)$$

#### Empirical Study

To analyze the performance of the proposed estimator compared to other estimators, three natural population data sets were considered. The description of the populations is given by Annexure.

Population I (Sukhatme & Sukhatme, 1970, p. 256):

y: Number of villages in the circles  
 x: A circle consisting more than five villages  
 $\bar{Y} = 3.360$   
 $\bar{X} = 0.1236$   
 $\rho = 0.766$   
 $C_y = 0.60400$   
 $C_x = 2.19012$

Population II (Cochran, 1977):

y: The number of persons per block  
 x: The number of rooms per block  
 $\bar{Y} = 101.1$

$\bar{X} = 58.80$   
 $\rho = 0.6500$   
 $C_y = 0.14450$   
 $C_x = 0.1281$

Population III (Kadilar & Singi, 2003):

y: Level of apple production  
 x: number of apple trees  
 $\bar{Y} = 625.37$   
 $\bar{X} = 13.93$   
 $\rho = 0.865$   
 $C_y = 1.866$   
 $C_x = 1.653,$

#### Results

Table 4.1 shows a significant gain in efficiency by using proposed estimator  $\hat{Y}_B$  ( $\hat{Y}_B^{(opt)}$ ) over the unbiased estimator  $\hat{Y}$ , the usual ratio estimator  $\hat{Y}_1$ , the product estimator  $\hat{Y}_2$ , the Sisodiya and Dwivedi (1981) estimator  $\hat{Y}_3$ , the Pandey and Dubey (1988) estimator  $\hat{Y}_4$ , the Singh and Tailor. (2003) estimators  $\hat{Y}_5$  and  $\hat{Y}_6$ , and the Kadilar and Singi (2006) estimator  $\hat{Y}_7$ .

Table 4.2 illustrates the wide range of  $\alpha$  in which a suggested estimator  $\hat{Y}_B$  or  $\hat{Y}_B^{(opt)}$  is more efficient than all estimators considered in this study; it shows that even if the scalar  $\alpha$  deviates from its optimum value ( $\alpha_{opt}$ ), the suggested estimator  $\hat{Y}_B^{(opt)}$  will yield better estimates than  $\hat{Y}$ ,  $\hat{Y}_1$ ,  $\hat{Y}_2$ ,  $\hat{Y}_3$ ,  $\hat{Y}_4$ ,  $\hat{Y}_5$ ,  $\hat{Y}_6$  and  $\hat{Y}_7$ . Therefore, the suggested estimator  $\hat{Y}_B^{(opt)}$  is recommended for use in practice.

IMPROVED ESTIMATION OF POPULATION MEAN

Table4.1: Percent Relative Efficiencies of  $\hat{Y}$ ,  $\hat{Y}_1$ ,  $\hat{Y}_2$ ,  $\hat{Y}_3$ ,  $\hat{Y}_4$ ,  $\hat{Y}_5$ ,  $\hat{Y}_6$ ,  $\hat{Y}_7$  and  $\hat{Y}_B$  or  $\hat{Y}_B^{(opt)}$  With Respect To  $\hat{Y}$

Estimators	$\hat{Y}$	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	$\hat{Y}_4$	$\hat{Y}_5$	$\hat{Y}_6$	$\hat{Y}_7$	$\hat{Y}_B^{(opt)}$
Population I	100.00	11.64	5.08	134.99	74.95	207.47	49.37	224.25	241.99
Population II	100.00	157.87	34.03	158.09	34.10	158.99	34.38	165.29	173.16
Population III	100.00	396.49	30.15	388.92	33.37	395.67	31.86	396.97	397.18

Table 4.2: Range of  $\alpha$  in Which  $\hat{Y}_B$  is Better than  $\hat{Y}$ ,  $\hat{Y}_1$ ,  $\hat{Y}_2$ ,  $\hat{Y}_3$ ,  $\hat{Y}_4$ ,  $\hat{Y}_5$ ,  $\hat{Y}_6$ ,  $\hat{Y}_7$

		Population		
		I	II	III
Range of $\alpha$ in which $\hat{Y}_B$ or $\hat{Y}_B^{(opt)}$ is Better Than	$\hat{Y}$	(0.5, 1.31)	(0.5, 1.30)	(0.5,1.51)
	$\hat{Y}_1$	(-0.413,0.165)	(0.638, 0.883)	(0.924 , 0.946)
	$\hat{Y}_2$	(-0.096, 0.22)	(-0.037,1.56)	(-0.017, 1.89)
	$\hat{Y}_3$	(0.041, 0.082)	(0.639, 0.883)	(0.895, 0.975)
	$\hat{Y}_4$	(0.027, 0.096)	(-0.036,1.56)	(0.034,1.84)
	$\hat{Y}_5$	(0.052, 0.057)	(0.512, 0.879)	(0.918, 0.964)
	$\hat{Y}_6$	(0.016, 0.087)	(-0.032,1.24)	(0.010, 1.88)
	$\hat{Y}_7$	(0.5, 0.809)	(0.5, 0.796)	(0.5, 0.013)
Optimum Value of $\alpha$ ( $\alpha_0$ )		(0.0617)	(0.7612)	(0.9350)

References

Sukhatme, P. V., & Sukhatme, B. V. (1970). *Sampling theory of surveys with applications*. Ames, IA: Iowa State University Press.

Das, A. K. (1988). *Contribution to the theory of sampling strategies based on auxiliary information*. Ph.D. Thesis, BCKV, West Bengal, India.

Pandey, B. N., & Dubey, V. (1988). Modified product estimator using coefficient of variation of auxiliary variable. *Assam Statistical Review*, 2, 64-66.

Singh, H. P., & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6, 555-560.

Sisodia, B. V. S., & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of Indian Society of Agricultural Statistics*, 33(1), 13-18.

Kadilar, C., & Cingi, H. (2006a). A new ratio estimator using correlation coefficient. *Inter Stat*, 1-11

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