

AN ALTERNATIVE RATIO-CUM-PRODUCT ESTIMATOR OF POPULATION MEAN USING A COEFFICIENT OF KURTOSIS FOR TWO AUXILIARY VARIATES

Rajesh Tailor^{1}, Med Ram Verma², and Balkishan Sharma³*

¹*School of Studies in Statistics, Vikram University, Ujjain, M.P., India*

Email: tailorraj@gmail.com

²*Division of Livestock Economics, Statistics and Information Technology, Indian Veterinary Research Institute (IVRI), Izatnagar, U.P., India*

³*Sri Aurobindo Institute of Medical Science, Indore, M. P., India*

ABSTRACT

An alternative ratio-cum-product estimator of population mean using the coefficient of kurtosis for two auxiliary variates has been proposed. The proposed estimator has been compared with a simple mean estimator, the usual ratio estimator, a product estimator, and estimators proposed by Singh (1967) and Singh et al. (2004). An empirical study is also carried out in support of the theoretical findings.

Key words: Population mean, Bias, Mean squared error, Coefficient of kurtosis

1 INTRODUCTION

The use of known parameters for auxiliary variates has played an important role in improving the efficiencies of estimators. Sisodiya and Dwivedi (1981) used the coefficient of variation for auxiliary variates. Later, Singh et al. (2004) used the coefficient of kurtosis for auxiliary variates. Upadhyaya and Singh (1999) derived ratio and product type estimators using both a coefficient of variation and a coefficient of kurtosis for auxiliary variates. Singh (1967) utilized information on two auxiliary variates x_1 and x_2 and suggested a ratio-cum-product estimator for population mean. This paper is an attempt to study the use of a coefficient of kurtosis ($\beta_2(x_1)$ and $\beta_2(x_2)$) for auxiliary variates in a ratio-cum product estimator.

Let $U = \{U_1, U_2, \dots, U_N\}$ be a finite population of N units. Suppose two auxiliary variates x_1 and x_2 are observed, along with study variate y , on U_i ($i = 1, 2, \dots, N$), where x_1 is positively and x_2 is negatively correlated with y . A simple random sample of size n is drawn by simple random sampling without replacement (SRSWOR) from the population U to estimate the population mean (\bar{Y}) of study character y when the population means

$\bar{X}_1 = \sum_{i=1}^N \frac{x_{1i}}{N}$ and $\bar{X}_2 = \sum_{i=1}^N \frac{x_{2i}}{N}$ of x_1 and x_2 , respectively, are known.

The usual ratio and product estimators for estimating the population mean \bar{Y} are given respectively by

$$\bar{y}_R = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right) \quad (1)$$

$$\bar{y}_P = \bar{y} \left(\frac{\bar{x}_2}{\bar{X}_2} \right) \quad (2)$$

Singh et al. (2004) defined a ratio and product type estimator using the coefficient of kurtosis ($\beta_2(x_1)$)

$$\bar{y}_{RB} = \bar{y} \left(\frac{\bar{X}_1 + \beta_2(x_1)}{\bar{x}_1 + \beta_2(x_1)} \right) \tag{3}$$

$$\bar{y}_{PB} = \bar{y} \left(\frac{\bar{x}_2 + \beta_2(x_2)}{\bar{X}_2 + \beta_2(x_2)} \right) \tag{4}$$

To estimate \bar{Y} , Singh (1967) suggested a ratio-cum-product estimator using information on two auxiliary variates x_1 and x_2 as

$$\hat{\bar{Y}}_1 = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right) \left(\frac{\bar{x}_2}{\bar{X}_2} \right) \tag{5}$$

To the first degree of approximation, the mean squared error (MSE) of the estimators $\bar{y}_R, \bar{y}_P, \bar{y}_{RB}, \bar{y}_{PB}$, and $\hat{\bar{Y}}_1$ are given respectively by

$$MSE(\bar{y}_R) = \theta \bar{Y}^2 [C_y^2 + C_{x_1}^2 - 2\rho_{yx_1} C_y C_{x_1}] \tag{6}$$

$$MSE(\bar{y}_P) = \theta \bar{Y}^2 [C_y^2 + C_{x_2}^2 + 2\rho_{yx_2} C_y C_{x_2}] \tag{7}$$

$$MSE(\bar{y}_{RB}) = \theta \bar{Y}^2 [C_y^2 + b_1^2 C_{x_1}^2 - 2\rho_{yx_1} b_1 C_y C_{x_1}] \tag{8}$$

$$MSE(\bar{y}_{PB}) = \theta \bar{Y}^2 [C_y^2 + b_2^2 C_{x_2}^2 + 2\rho_{yx_2} b_2 C_y C_{x_2}] \tag{9}$$

$$MSE(\hat{\bar{Y}}_1) = \theta \bar{Y}^2 [C_y^2 + C_{x_1}^2 (1 - 2K_{yx_1}) + C_{x_2}^2 \{1 + 2(K_{yx_2} - K_{x_1x_2})\}] \tag{10}$$

Where

$$K_{yx_1} = \rho_{yx_1} \left(\frac{C_y}{C_{x_1}} \right), K_{yx_2} = \rho_{yx_2} \left(\frac{C_y}{C_{x_2}} \right), K_{x_1x_2} = \rho_{x_1x_2} \left(\frac{C_{x_1}}{C_{x_2}} \right), C_y = \frac{S_y}{\bar{Y}}$$

$$\theta = \left(\frac{1}{n} - \frac{1}{N} \right), C_{x_i} = \frac{S_{x_i}}{\bar{X}_i}, \rho_{yx_i} = \frac{S_{yx_i}}{(S_y S_{x_i})}, b_1 = \frac{\bar{X}}{\bar{X} + \beta_2(x_1)}, b_2 = \frac{\bar{X}}{\bar{X} + \beta_2(x_2)}$$

$$S_y^2 = \frac{\sum_{j=1}^N (y_j - \bar{Y})^2}{(N-1)}, S_{x_i}^2 = \frac{\sum_{j=1}^N (x_{ij} - \bar{X}_i)^2}{(N-1)} \text{ and } S_{yx_i}^2 = \frac{\sum_{j=1}^N (y_j - \bar{Y})(x_{ij} - \bar{X}_i)}{(N-1)}, \text{ where } (i=1,2).$$

2 PROPOSED ESTIMATOR

Assuming that the information on the coefficient of kurtosis ($\beta_2(x_1)$ & $\beta_2(x_2)$) for auxiliary variates x_1 and x_2 is available, the proposed estimator of \bar{Y} is

$$\hat{\bar{Y}}_2 = \bar{y} \left(\frac{\bar{X}_1 + \beta_2(x_1)}{\bar{x}_1 + \beta_2(x_1)} \right) \left(\frac{\bar{x}_2 + \beta_2(x_2)}{\bar{X}_2 + \beta_2(x_2)} \right) \tag{11}$$

When information on the second auxiliary variable x_2 is not available (or equivalently, the variable x_2 takes only a constant value, i.e., $x_{2i} = a$ (constant); $i = 1, 2, \dots, N$), the estimator $\hat{\bar{Y}}_2$ reduces to \bar{y}_{RB} as suggested by Singh et al. (2004). On the other hand, if the information on the auxiliary variable x_1 is not available (or equivalently the

variable x_i takes only a constant value, i.e., $x_{1i} = a^*$ (constant); $i = 1, 2, \dots, N$), the estimator turns out to be the estimator \bar{y}_{PB} , a product version of \bar{y}_{RB} .

To obtain the bias and mean squared error of the proposed estimators, we assume that

$$\bar{y} = \bar{Y}(1 + e_0), \bar{x}_1 = \bar{X}_1(1 + e_1) \text{ and } \bar{x}_2 = \bar{X}_2(1 + e_2) \text{ such that}$$

$$E(e_0) = E(e_1) = E(e_2) = 0, \text{ and}$$

$$E(e_0^2) = \theta C_y^2, E(e_1^2) = \theta C_{x_1}^2, E(e_2^2) = \theta C_{x_2}^2$$

$$E(e_0 e_1) = \theta \rho_{yx_1} C_y C_{x_1}, E(e_0 e_2) = \theta \rho_{yx_2} C_y C_{x_2} \text{ and } E(e_1 e_2) = \theta \rho_{x_1 x_2} C_{x_1} C_{x_2}$$

Expressing \hat{Y}_2 in terms of e_i 's, we get

$$\hat{Y}_2 = \bar{Y}(1 + e_0)(1 + b_1 e_1)^{-1}(1 + b_2 e_2) \tag{12}$$

The bias and mean squared error of \hat{Y}_2 are

$$B(\hat{Y}_2) = \theta \bar{Y} [b_1 C_{x_1}^2 (b_1 - K_{yx_1}) + b_2 C_{x_2}^2 (K_{yx_2} - b_1 K_{x_1 x_2})] \tag{13}$$

$$MSE(\hat{Y}_2) = \theta \bar{Y}^2 [C_y^2 + b_1 C_{x_1}^2 (b_1 - 2K_{yx_1}) + [b_2 C_{x_2}^2 \{b_2 + 2(K_{yx_2} - b_1 K_{x_1 x_2})\}]] \tag{14}$$

3 EFFICIENCY COMPARISON

The variance of sample mean \bar{y} in simple random sampling without replacement (SRSWOR) is

$$V(\bar{y}) = \theta S_y^2 \tag{15}$$

From equations (6) to (9), (14), and (15) we have

(i) $MSE(\hat{Y}_2) < MSE(\bar{y})$ if

$$K_{yx_1} > \left(\frac{b_1}{2}\right) \text{ and } K_{yx_2} < \left(b_1 K_{x_1 x_2} - \frac{b_2}{2}\right) \tag{16}$$

(ii) $MSE(\hat{Y}_2) < MSE(\bar{y}_R)$ if

$$K_{yx_1} < \left(\frac{1+b_1}{2}\right) \text{ and } K_{yx_2} < \left(b_1 K_{x_1 x_2} - \frac{b_2}{2}\right) \tag{17}$$

(iii) $MSE(\hat{Y}_2) < MSE(\bar{y}_P)$ if

$$K_{yx_1} > \left(\frac{b_1}{2} - b_2 K_{x_2 x_1}\right) \text{ and } K_{yx_2} > -\left(\frac{(1+b_2)}{2}\right) \tag{18}$$

(iv) $MSE(\hat{Y}_2) < MSE(\hat{Y}_1)$ if

$$K_{yx_1} < \left(\frac{(1+b_1)}{2}\right) \text{ and } K_{yx_2} > \left(\frac{K_{x_1 x_2} (1-b_1 b_2)}{(1-b_1)} - \frac{(1+b_2)}{2}\right) \tag{19}$$

4 A FAMILY OF UNBIASED ESTIMATORS USING THE JACKKNIFE TECHNIQUE

Suppose a simple random sample of size $n = gm$ is drawn without replacement and split at random into g sub-samples, each of size m . Then the Jack-knife type ratio-cum-product estimator for population mean \bar{Y} , using \hat{Y}_2 is given as

$$\hat{Y}_{2J} = \frac{1}{g} \sum_{j=1}^g \bar{y}'_j \left(\frac{\bar{X}_1 + \beta_2(x_1)}{\bar{x}'_{1j} + \beta_2(x_1)} \right) \left(\frac{\bar{x}'_{2j} + \beta_2(x_2)}{\bar{X}_2 + \beta_2(x_2)} \right) \tag{20}$$

where $\bar{y}'_j = (n\bar{y} - m\bar{y}_j)/(n - m)$ and $\bar{x}'_{ij} = (n\bar{x}_i - m\bar{x}_{ij})/(n - m)$, $i=1,2$ are the sample means based on a sample of $(n-m)$ units obtained by omitting the j^{th} group and \bar{y}_j and \bar{x}_{ij} ($i=1,2; j=1,2,\dots,g$) are the sample means based on the j^{th} sub samples of size $m=n/g$.

The bias of \hat{Y}_{2J} , up to the first degree of the approximation bias of \hat{Y}_2 , is obtained as

$$B(\hat{Y}_{2J}) = \frac{(N - n + m)}{N(n - m)} \bar{Y} \left[b_1 C_{x_1}^2 (b_1 - K_{yx_1}) + b_2 C_{x_2}^2 (K_{yx_2} - b_1 K_{x_1x_2}) \right]. \tag{21}$$

From (13) and (21) we have

$$\frac{B(\hat{Y}_2)}{B(\hat{Y}_{2J})} = \frac{(N - n)(n - m)}{n(N - n + m)} \tag{22}$$

Upon simplifying (22), we get a general family of almost unbiased ratio-cum-product estimators of \bar{Y} as

$$\hat{Y}_{2u} = \left[\bar{y} \{1 - b^* (1 - \delta^*)\} + b^* \hat{Y}_2 - b^* \delta^* \hat{Y}_{2J} \right]. \tag{23}$$

Remark 4.1. For $b^* = 0$, \hat{Y}_{2u} yields the usual unbiased estimator \bar{y} , while $b^* = (1 - \delta^*)^{-1}$ gives an almost unbiased estimator for \bar{Y} as

$$\hat{Y}_{2u}^* = \frac{(N - n + m)}{N} g \bar{y} \left(\frac{\bar{X}_1 + \beta_2(x_1)}{\bar{x}_1 + \beta_2(x_1)} \right) \left(\frac{\bar{x}_2 + \beta_2(x_2)}{\bar{X}_2 + \beta_2(x_2)} \right) - \frac{(N - n)(g - 1)}{Ng} \sum_{j=1}^g \bar{y}'_j \left(\frac{\bar{X}_1 + \beta_2(x_1)}{\bar{x}'_{1j} + \beta_2(x_1)} \right) \left(\frac{\bar{x}'_{2j} + \beta_2(x_2)}{\bar{X}_2 + \beta_2(x_2)} \right) \tag{24}$$

This is the Jack-knifed version of the proposed estimator \hat{Y}_2 .

5 AN OPTIMUM ESTIMATOR IN FAMILY \hat{Y}_{2u}

The family of the almost unbiased estimator \hat{Y}_{2u} in (23) can be expressed as

$$\hat{Y}_{2u} = \bar{y} - \lambda^* \bar{y}_1, \tag{25}$$

where $\bar{y}_1 = [(1 - \delta^*)\bar{y} - \bar{y}_2]$ and $\bar{y}_2 = \hat{Y}_2 - \delta^* \hat{Y}_{2J}$.

The variance of \hat{Y}_{2u} is given by

$$V(\hat{Y}_{2u}) = V(\bar{y}) + b^{*2} V(\bar{y}_1) - 2b^* Cov(\bar{y}, \bar{y}_1), \quad (26)$$

which is minimized for

$$b^* = Cov(\bar{y}, \bar{y}_1) / V(\bar{y}_1). \quad (27)$$

Substitution of (27) in (26) yields the minimum variance of \hat{Y}_{2u} as

$$\min.V(\hat{Y}_{2u}) = V(\bar{y}) - \frac{\{Cov(\bar{y}, \bar{y}_1)\}^2}{V(\bar{y}_1)} = V(\bar{y})(1 - \rho_{01}^2) \quad , \quad (28)$$

where ρ_{01} is the correlation coefficient between \bar{y} and \bar{y}_1 .

From (28) it is clear that $\min.V(\hat{Y}_{2u}) < V(\bar{y})$.

To obtain the explicit expression of the variance of \hat{Y}_{2u} , we write the following results up to terms of order n^{-1} , as

$$MSE(\hat{Y}_{2J}) = Cov(\hat{Y}_2, \hat{Y}_{2J}) = MSE(\hat{Y}_2) \quad , \quad (29)$$

and

$$Cov(\bar{y}, \hat{Y}_2) = Cov(\bar{y}, \hat{Y}_{2J}) = \theta \bar{Y}^2 [C_y^2 - b_1 \rho_{yx_1} C_y C_{x_1} + b_2 \rho_{yx_2} C_y C_{x_2}], \quad (30)$$

where $MSE(\hat{Y}_2)$ is given by (14).

Using (14), (15), and (30) in (26), the variance of \hat{Y}_{2u} up to the terms of order n^{-1} is given as

$$V(\hat{Y}_{2u}) = \theta \bar{Y}^2 [C_y^2 + b^{*2} (1 - \delta^*)^2 (b_1^2 C_{x_1}^2 + b_2^2 C_{x_2}^2 - 2\rho_{x_1 x_2} C_{x_1} C_{x_2} b_1 b_2) - 2b^* (1 - \delta^*) (b_1 \rho_{yx_1} C_y C_{x_1} - b_2 \rho_{yx_2} C_y C_{x_2})] \quad , \quad (31)$$

which is minimized for

$$b^* = \frac{(b_1 \rho_{yx_1} C_y C_{x_1} - b_2 \rho_{yx_2} C_y C_{x_2})}{(1 - \delta^*) (b_1^2 C_{x_1}^2 + b_2^2 C_{x_2}^2 - 2b_1 b_2 \rho_{x_1 x_2} C_{x_1} C_{x_2})} = b_{opt}^* \quad (32)$$

Substitution of the value of b_{opt}^* in \hat{Y}_{2u} yields the optimum estimator $\hat{Y}_{2u(opt)}$ (say). Thus the resulting minimum

variance of \hat{Y}_{2u} is given by

$$\min.V(\hat{Y}_{2u}) = \theta \bar{Y}^2 C_y^2 \left[1 - \frac{(b_1 \rho_{yx_1} C_{x_1} - b_2 \rho_{yx_2} C_{x_2})^2}{(b_1^2 C_{x_1}^2 + b_2^2 C_{x_2}^2 - 2b_1 b_2 \rho_{x_1 x_2} C_{x_1} C_{x_2})} \right] = V(\hat{Y}_{2u(opt)}) \quad (33)$$

The optimum value b_{opt}^* of b^* can be obtained quite accurately through past data or experience.

6 EMPIRICAL STUDY

To observe the relative performance of different estimators of \bar{Y} , a natural population data set is considered.

Population [Source: Steel and Torrie (1960, p.282)]

y : Log of leaf burn in sec., x_1 : Potassium percentage, x_2 : Chlorine percentage.

The required population parameters are

$$\bar{Y} = 0.6860, \quad C_y = 0.4803, \quad \rho_{yx_1} = 0.1794, \quad N=30,$$

$$\bar{X}_1 = 4.6537, \quad C_{x_1} = 0.2295, \quad \rho_{yx_2} = -0.4996, \quad n=6,$$

$$\bar{X}_2 = 0.8077, \quad C_{x_2} = 0.7493, \quad \rho_{x_1x_2} = 0.4074.$$

Table 1. Percent relative efficiencies of different estimators of \bar{Y} with respect to \bar{y}

Estimators	\bar{y}	\bar{y}_R	\bar{y}_P	\bar{y}_{Rb}	\bar{y}_{Pb}	\hat{Y}_1	\hat{Y}_2	$\hat{Y}_2 (\hat{Y}_2^{opt})$
PREs	100.00	94.62	53.33	100.03	132.37	75.50	169.87	173.81

7 RESULT AND DISCUSSION

The proposed estimator \hat{Y}_2 would be more efficient than other estimators if its mean squared error is less than the mean squared error of other estimators. Under conditions (16), (17), (18), and (19), the mean squared error of the proposed estimator would be less than the mean squared error of the sample mean estimator \bar{y} , ratio estimator \bar{y}_R , product estimator \bar{y}_P , and Singh (1967) estimator \hat{Y}_1 , respectively. Thus under these conditions, the proposed estimator would be more efficient.

Table 1 reveals that the suggested estimators $\hat{Y}_2 (\hat{Y}_2^{opt})$ are more efficient than the usual unbiased estimator \bar{y} , ratio estimator \bar{y}_R , product estimator \bar{y}_P , ratio and product type estimators estimator \bar{y}_{Rb} and \bar{y}_{Pb} proposed by Singh et al. (2004), and the ratio-cum-product estimator \hat{Y}_1 suggested by Singh (1967) with considerable gain in efficiency. Thus, if the coefficients of kurtosis ($\beta_2(x_1)$ and $\beta_2(x_2)$) are known, the suggested estimator is recommended for use in practice.

8 ACKNOWLEDGEMENT

The authors are thankful to the referee for his valuable suggestions regarding improvement of the paper.

9 REFERENCES

Quenouille, M.H. (1956) Notes on bias in estimation. *Biometrika*, 43, 353-360.

Singh, H.P., Tailor, R., Tailor, R., & Kakran, M.S. (2004) An Improved Estimation of Population Mean Using Power Transformation. *Jour. Ind. Soc. Agri. Stat.*, 58 (2), 223-230

Singh, M.P. (1967) Ratio-cum-product method of estimation, *Metrika*, 12, 34-42.

Sisodia, B.V.S. & Dwivedi, V.K. (1981) A modified ratio estimator using coefficient of variation of auxiliary variable. *Jour. Ind. Soc. Agri. Stat.*, 33(1), 13-18.

Steel, R.G.D. & Torrie J.H. (1960) Principles and Procedures of Statistics. Mc Graw Hill Book Co.

Upadhyaya L.N. & Singh, H.P. (1999) Use of transformed auxiliary variable in estimating the finite population mean, *Bio. Jour.* 41(5) 627-636.

(Article history: Received 14 July 2009, Accepted 1 June 2010, Available online 29 June 2010)