Jour. Ind. Soc. Ag. Statistics
56(3), 2003 : 267-275

On the Improvement of Product Method of Estimation in Sample Surveys

G.N. Singh
Indian School of Mines, Dhanbad-826 004
(Received : July, 2002)

SUMMARY

For estimating a finite population mean under product method of estimation, simple linear transformations using the known coefficient of skewness, coefficient of kurtosis and standard deviation of the auxiliary variable have been considered. The suggested transformations provide efficient product estimators with less absolute bias than conventional product and other product type estimators.

Key words : Product method, Coefficient of variation (CV), Coefficient of skewness (CS), Coefficient of kurtosis (CK), Bias, Mean square error.

1. Introduction

Product method of estimation is well-known technique for estimating the population mean of a study character when population mean of an auxiliary character is known and it is negatively correlated with study character.

To find more precise estimates, Searls [5] used CV of study character at estimation stage. In practice this CV is seldom known. Motivated by Searls [5] work, Sisodia and Dwivedi [9] used the known CV of the auxiliary character for estimating population mean of a study character in ratio method of estimation. Following the work of Sisodia and Dwivedi [9], Pandey and Dubey [4] proposed a modified product estimator for population mean of a study character using known CV of an auxiliary character. Recently Upadhyaya and Singh [12] proposed new product estimators using known CV and coefficient of kurtosis (CK) of an auxiliary character. All these authors have used known CV and CK of an auxiliary character in additive form to sample and population means of the same character. It could be noticed that CV and CK are unit free constants therefore their additions may not be justified. Further, if CV and population mean of an auxiliary character are known, standard deviation (SD) of the same auxiliary character is automatically known and use of standard deviation in additive form is more justified.

Motivated by the above points, in this work, an attempt has been made to utilize the information on known standard deviation (SD) of an auxiliary character of skewness (CS) and coefficient of kurtosis (CK) of an auxiliary character through a most suitable transformation for estimating population mean of a study character. Three new product estimators for population mean of study character have been proposed. The proposed estimators have been compared with the simple mean, conventional product estimator and the product estimators proposed by Pandey and Dubey [4] and Upadhyaya and Singh [12].

The performances of the proposed estimators have been supported with a numerical illustration.

2. *Proposed Estimators*

Let $U = (u_1, u_2, ..., u_N)$ be the finite population of N units, y and x be the character under study and auxiliary character, respectively. It is assumed that y and x are highly negatively correlated. Let $y_k > 0$ and $x_k > 0$ be the values of y and x for k-th (k = 1, 2, ..., N) unit in the population. From the population U, a simple random sample of size n is drawn without replacement. Let $(\bar{Y}, \bar{X})$ and $(\bar{y}, \bar{x})$ be the population means and sample means of the respective variates y and x, respectively. The conventional product estimator for $\bar{Y}$ is defined as

$$\bar{y}_p = \frac{\bar{y}\,\bar{x}}{\bar{X}} \qquad (2.1)$$

Utilizing the known CV of x, i.e., C_x, Pandey and Dubey [4] considered the modified product estimator of $\bar{Y}$ as

$$\bar{y}_{mp} = \bar{y}\left[\frac{\bar{x} + C_x}{\bar{X} + C_x}\right] \qquad (2.2)$$

In many situations the values of an auxiliary variable may be available for each unit in the population, for instance, see Das and Tripathi [1]. In such situations knowledge on $\bar{X}$, C_x, $\beta_1(x)$ (coefficient of skewness), $\beta_2(x)$ (coefficient of kurtosis) and possibly on some other parameters may be utilized. Regarding the availability of information on $C_x, \beta_1(x)$ and $\beta_2(x)$, the researchers may refer to Searls [5], Sen [7], Singh *et al.* [8] and Searls and Intarpanich [6].

Using C_x and $\beta_2(x)$ of the auxiliary variable x, Upadhyaya and Singh [12] suggested the two more modified product estimators which are reproduced below

$$\bar{y}_{us1} = \bar{y}\left[\frac{\beta_2(x)\bar{x} + C_x}{\beta_2(x)\bar{X} + C_x}\right] \tag{2.3}$$

and

$$\bar{y}_{us2} = \bar{y}\left[\frac{C_x\bar{x} + \beta_2(x)}{C_x\bar{X} + \beta_2(x)}\right] \tag{2.4}$$

If the population mean and CV of an auxiliary character are known, the standard deviation σ_x is automatically known and it is more meaningful to use σ_x in addition to C_x see Srivastava and Jhajj [10]. Further, C_x, $\beta_1(x)$ and $\beta_2(x)$ are the unit free constants, their use in additive form is not much justified. Motivated with the above arguments and utilizing the known values of σ_x, $\beta_1(x)$ and $\beta_2(x)$, we suggest the following three more reasonable transformations for x and corresponding estimators for $\bar{Y}$, which are as follows.

Let $z_{ik} = \alpha_i x_k + \sigma_x$, (i = 1, 2, 3 and k = 1, 2, ..., N) so that $\bar{z}_i = \alpha_i\bar{x} + \sigma_x$ is the sample mean of the transformed variable z and $\bar{Z}_i = \alpha_i\bar{X} + \sigma_x$ is the corresponding population mean, where $\alpha_1 = 1, \alpha_2 = \beta_1(x)$ and $\alpha_3 = \beta_2(x)$ are the known values of α_i's for i = 1, 2, 3. Subsequently, the following three new modified product estimators of $\bar{Y}$ are considered as

$$T_{p1} = \bar{y}\left[\frac{\bar{x} + \sigma_x}{\bar{X} + \sigma_x}\right] = \frac{\bar{y}\bar{z}_1}{\bar{Z}_1}, \text{ for } i = 1, \alpha_1 = 1 \tag{2.5}$$

$$T_{p2} = \bar{y}\left[\frac{\beta_1(x)\bar{x} + \sigma_x}{\beta_1(x)\bar{X} + \sigma_x}\right] = \frac{\bar{y}\bar{z}_2}{\bar{Z}_2}, \text{ for } i = 2, \alpha_2 = \beta_1(x) \tag{2.6}$$

and

$$T_{p3} = \bar{y}\left[\frac{\beta_2(x)\bar{x} + \sigma_x}{\beta_2(x)\bar{X} + \sigma_x}\right] = \frac{\bar{y}\bar{z}_3}{\bar{Z}_3}, \text{ for } i = 3, \alpha_3 = \beta_2(x) \tag{2.7}$$

Estimators proposed in (2.5) to (2.7) can be re-written as a sequence of new modified product estimators in the following form

$$T_{pi} = \bar{y}\left[\frac{\alpha_i\bar{x} + \sigma_x}{\alpha_i\bar{X} + \sigma_x}\right] = \frac{\bar{y}\bar{z}_i}{\bar{Z}_i}, (i = 1, 2, 3) \tag{2.8}$$

3. Bias and Mean Square Error (MSE) of the Proposed Sequence of Estimators T_{pi} (i = 1, 2, 3)

The proposed sequence of estimators T_{pi}, (i = 1, 2, 3) is biased for $\overline{Y}$. The exact bias and mean square error (m.s.e) up to the order $0(n^{-1})$ have been derived under the following approximations

$\overline{y} = \overline{Y}(1 + e_1)$, $\overline{x} = \overline{X}(1 + e_2)$ with $E(e_j) = 0$, for j = 1, 2. The proposed sequence of estimators T_{pi}, (i = 1, 2, 3) then becomes

$$T_{pi} = \overline{Y}(1 + e_1)(1 + \varphi_i e_2), (i = 1, 2, 3) \tag{3.1}$$

where $\varphi_1 = \dfrac{\overline{X}}{\overline{X} + \sigma_x}$, $\varphi_2 = \dfrac{\beta_1(x)\overline{X}}{\beta_1(x)\overline{X} + \alpha_x}$ and $\varphi_3 = \dfrac{\beta_2(x)\overline{X}}{\beta_2(x)\overline{X} + \sigma_x}$

For i = 1, 2, 3; $\varphi_i < 1$, now expanding the terms of (3.1) and taking expectations, we have the following results

Theorem 1. The exact bias B(.) and mean square error (m.s.e.) M(.) of the proposed sequence of estimators T_{pi}, (i = 1, 2, 3) to the terms of order $0(n^{-1})$ are given by

$$B(T_{pi}) = E(T_{pi} - \overline{Y}) = \varphi_i f_1 \overline{Y} \rho_{yx} C_y C_x \tag{3.2}$$

and

$$M(T_{pi}) = E(T_{pi} - \overline{Y})^2 = \overline{Y}^2 f_1 (C_y^2 + \varphi_i^2 C_x^2 + 2\varphi_i \rho_{yx} C_y C_x) \tag{3.3}$$

where ρ_{yx} is the population coefficient of correlation between y and x, C_y is the coefficient of variation of the variable y and $f_1 = \left(\dfrac{1}{n} - \dfrac{1}{N}\right)$.

Corollary 1. Biases and m.s.e.'s of T_{p1}, T_{p2} and T_{p3} defined in (2.5) to (2.7) can be obtained by substituting the values of φ_i, (i = 1, 2, 3) in (3.2) and (3.3), respectively.

4. Efficiency Comparisons of T_{pi} (i = 1, 2, 3)

In this section, the conditions for which the proposed sequence of estimators are better than $\overline{y}$, $\overline{y}_p$, $\overline{y}_{mp}$, $\overline{y}_{us1}$ and $\overline{y}_{us2}$ have been obtained. The exact variance of sample mean estimator $\overline{y}$, expressions for exact biases and the m.s.e's of these estimators up to the order $0(n^{-1})$ are given below

$$B(\overline{y}) = 0 \tag{4.1}$$

$$B(\bar{y}_p) = \bar{Y} f_1 \rho_{yx} C_y C_x \tag{4.2}$$

$$B(\bar{y}_{mp}) = \bar{Y} f_1 P \rho_{yx} C_y C_x \tag{4.3}$$

$$B(\bar{y}_{us1}) = \bar{Y} f_1 Q \rho_{yx} C_y C_x \tag{4.4}$$

$$B(\bar{y}_{us2}) = \bar{Y} f_1 R \rho_{yx} C_y C_x \tag{4.5}$$

$$V(\bar{y}) = \bar{Y}^2 f_1 C_y^2 \tag{4.6}$$

$$M(\bar{y}_p) = \bar{Y}^2 f_1 (C_y^2 + C_x^2 + 2\rho_{yx} C_y C_x) \tag{4.7}$$

$$M(\bar{y}_{mp}) = \bar{Y}^2 f_1 (C_y^2 + P^2 C_x^2 + 2P\rho_{yx} C_y C_x) \tag{4.8}$$

$$M(\bar{y}_{us1}) = \bar{Y}^2 f_1 (C_y^2 + Q^2 C_x^2 + 2Q\rho_{yx} C_y C_x) \tag{4.9}$$

$$M(\bar{y}_{us2}) = \bar{Y}^2 f_1 (C_y^2 + R^2 C_x^2 + 2R\rho_{yx} C_y C_x) \tag{4.10}$$

where $P = \dfrac{\bar{X}}{\bar{X} + C_x}$, $Q = \dfrac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X} + C_x}$, and $R = \dfrac{C_x \bar{X}}{C_x \bar{X} + \beta_2(x)}$

(i) Comparison of T_{pi}, $(i = 1, 2, 3)$ with $\bar{y}$ and $\bar{y}_p$

It is obvious from (3.3) and (4.6) that T_{pi} is preferable over simple mean estimator $\bar{y}$ if

$$\rho_{yx} < -\frac{\varphi_i}{2}\frac{C_x}{C_y} \tag{4.11}$$

Here, the range of ρ_{yx} is widened as compared to the range of ρ_{yx} for which $\bar{y}_p$ is more efficient than $\bar{y}$. It shows that T_{pi} provides more precise estimates even for some low values of ρ_{yx}.

Since $\varphi_i < 1$ for $i = 1, 2, 3$ it follows from (3.3) and (4.7) that the sequence of estimators T_{pi} is more efficient than $\bar{y}_p$ if

$$\rho_{yx} > -\frac{(\varphi_i + 1)}{2}\frac{C_x}{C_y} \tag{4.12}$$

subsequently the results in (4.11) and (4.12) are combined in the following theorem

Theorem 2. The range of ρ_{yx} for which T_{pi}, $(i = 1, 2, 3)$ is better than both $\bar{y}$ and $\bar{y}_p$ is

$$-\frac{(\varphi_i+1)}{2}\frac{C_x}{C_y} < \rho_{yx} < -\frac{\varphi_i}{2}\frac{C_x}{C_y} \qquad (4.13)$$

Corollary 2. Suitable ranges of ρ_{yx} for the estimators, T_{p1}, T_{p2} and T_{p3} can be derived by substituting the respective values of φ_i, (i = 1, 2, 3) in (4.13).

(ii) Comparison of T_{pi}, (i = 1, 2, 3) with $\bar{y}_{mp}$, $\bar{y}_{us1}$ and $\bar{y}_{us2}$

From (3.3), (4.8), (4.9) and (4.10) it could be concluded that

(a) T_{pi} is better than $\bar{y}_{mp}$ if

$$\rho_{yx} < -\frac{(\varphi_i+P)}{2}\frac{C_x}{C_y}, \quad (i=1,2,3) \qquad (4.14)$$

(b) T_{pi} is more precise than $\bar{y}_{us1}$ if

$$\rho_{yx} < -\frac{(\varphi_i+Q)}{2}\frac{C_x}{C_y}, \quad (i=1,2,3) \qquad (4.15)$$

and

(c) T_{pi} dominates $\bar{y}_{us2}$ if

$$\rho_{yx} < -\frac{(\varphi_i+R)}{2}\frac{C_x}{C_y}, \quad (i=1,2,3) \qquad (4.16)$$

Remark. When the study character y and auxiliary character x are positively correlated, we may consider the following sequence of new modified ratio estimators as

$$T_{ri} = \bar{y}\left[\frac{\alpha_i\bar{X}+\sigma_x}{\alpha_i\bar{x}+\sigma_x}\right] = \frac{\bar{y}\bar{Z}_i}{\bar{z}_i}, \quad (i=1,2,3) \qquad (4.17)$$

5. *Unbiased Versions of the Suggested Sequence of Estimators T_{pi} (i=1, 2, 3)*

We find the unbiased versions of the suggested sequence of estimators T_{pi} using two well-known techniques: (i) Interpenetrating sub-samples design and (ii) Jack-knife technique.

(i) Interpenetrating Sub-samples Design

Let the sample in the form of n independent interpenetrating sub-samples be drawn. Let y_k and x_k be unbiased estimates of the population totals Y and X respectively based on the k-th independent interpenetrating sub-sample, k = 1, 2, ..., n. We now consider two different sequences of product type estimators

$$T_{pi} = \frac{\bar{y}\,\bar{z}_i}{\bar{Z}_i}, \ (i = 1, 2, 3) \tag{5.1}$$

and $$T^{*}_{pi} = \sum_{k=1}^{n} y_k z_{ik} / (n\bar{Z}_i), \ z_{ik} = \alpha_i x_k + \sigma_x, \ (i = 1, 2, 3) \tag{5.2}$$

Following Murthy [3], we can show that $B(T^{*}_{pi}) = nB(T_{pi})$ and hence that

$$T_{upi} = \frac{nT_{pi} - T^{*}_{pi}}{(n-1)} \text{ is unbiased for } \bar{Y} \tag{5.3}$$

The properties of new unbiased sequence of product estimators T_{upi}, (i = 1, 2, 3) can be studied on the lines of Murthy and Nanjamma [2].

(ii) Jack-knife Technique

We may now take n = 2m and split the sample at random into two sub-samples of m units each. Let $\bar{y}_k, \bar{x}_k (k = 1, 2)$ be unbiased estimators of population means $\bar{Y}$ and $\bar{X}$ based on the sub-samples and $\bar{y}, \bar{x}$ the means based on the entire sample.

Take $\bar{z}_{ik} = \alpha_i \bar{x}_k + \sigma_x$ and $\bar{Z}_i = \alpha_i \bar{X} + \sigma_x$, (i = 1, 2, 3 and k = 1, 2)

Thus the unbiased version of sequence of product estimators is given by

$$T^{(u)}_{pi} = \frac{(2N-n)}{N} T_{pi} - \frac{(N-n)}{2N} \left\{ T^{(1)}_{pi} + T^{(2)}_{pi} \right\}, \ (i = 1, 2, 3) \tag{5.4}$$

where $T^{(k)}_{pi} = \frac{\bar{y}_k\,\bar{z}_{ik}}{\bar{Z}_i}$, (i = 1, 2, 3 and k = 1, 2) and T_{pi} defined in (2.8).

The variance expression of the sequence of unbiased product estimators $T^{(u)}_{pi}$, (i = 1, 2, 3) can be derived on the lines similar to those of Sukhatme and Sukhatme [11] (pp 161-165). It is interesting to note that the variance expression of $T^{(u)}_{pi}$ and the m.s.e. expression of T_{pi} are equal up to the first order of approximation. Hence one can prefer the sequence of unbiased product estimator $T^{(u)}_{pi}$ as compared to T_{pi} as $T^{(u)}_{pi}$ is unbiased.

6. Empirical Study

We consider an artificial data used by Pandey and Dubey [4]. The population constants are as follows

$N = 20, n = 8, \overline{Y} = 19.55, \overline{X} = 18.8, C_x^2 = 0.1555, C_y^2 = 0.1262, \rho_{yx} = -0.9199$

$P = 0.9795, Q = 0.9932, R = 0.7077, \varphi_1 = 0.7172, \varphi_2 = 0.5812, \varphi_3 = 0.8859$

$\beta_1(x) = 0.5473, \beta_2(x) = 3.0613$ and $f_1 = 0.075$

The percent relative bias and percent relative efficiencies of the estimators with respect to simple mean estimator are shown in Table 1.

Table 1

Estimator	Percent Relative Absolute Bias	Percent Relative Efficiency
$\overline{y}$	0	100
$\overline{y}_p$	0.9660	525.83
$\overline{y}_{mp}$	0.9462	548.70
$\overline{y}_{us1}$	0.9594	532.49
$\overline{y}_{us2}$	0.6836	581.57
T_{p1}	0.6928	589.72
T_{p2}	0.5614	436.68
T_{p3}	0.8558	634.17

From Table 1, we observe that T_{p1} and T_{p3} are preferable over the product and other modified product estimators. Estimator T_{p2} is not performing much better in terms of m.s.e., but its absolute bias is least among the list of estimators shown in the table. It could be observed that T_{p2} will perform well for the highly skewed populations.

7. Conclusions

The transformations which are considered in this work are more reasonable and from the above results, we can conclude that the estimators which are proposed in this work are either better in terms of bias or in terms of m.s.e. criterion. While sometimes it is preferable in terms of both the criterions. Therefore, survey statisticians may recommend the use of these estimators.

ACKNOWLEDGEMENTS

Author is thankful to the referees for their valuable and inspiring comments.

REFERENCES

[1] Das, A.K. and Tripathi, T.P. (1981). A class of sampling strategies for population mean using information on mean and variance of an auxiliary character. *Proc. of the ISI Golden Jubilee International Conference on Statistics: Applications and New Directions*. ISI Calcutta, 16-19 Dec., 174-181.

[2] Murthy, M.N. and Nanjamma (1959). Almost unbiased ratio estimates based on interpenetrating sub-sample estimates. *Sankhya*, **21**, 381-392.

[3] Murthy, M.N. (1964). Product method of estimation. *Sankhya*, **26A**, 294-307.

[4] Pandey, B.N. and Dubey, Vyas (1988). Modified product estimator using coefficient of variation of auxiliary variate. *Assam Stat. Rev.*, **2(2)**, 64-66.

[5] Searls, D.T. (1964). The utilization of known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.*, **59**, 1125-1126.

[6] Searls, D.T. and Intarpanich, P. (1990). A note on an estimator for the variance that utilizes the kurtosis. *Amer. Stat.*, **44**(4), 295-296.

[7] Sen, A.R. (1978). Estimation of the population mean when the coefficient of variation is known. *Comm. Stat.–Theory Methods*, **A7**, 657-672.

[8] 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.

[9] Sisodia, B.V.S. and Dwivedi, V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Jour. Ind. Soc. Agril. Stat.*, **33(2)**, 13-18.

[10] Srivastava, S.K. and Jhajj, H.S. (1980). A class of estimators using auxiliary information for estimating population variance. *Sankhya*, **C42**, 87-96.

[11] Sukhatme, P.V. and Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Applications*. 2nd Edition. Iowa State University Press, Ames, USA.

[12] Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biom. J.*, **41(5)**, 627-636.