JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A MODIFIED PRODUCT-TYPE ESTIMATOR FOR ESTIMATING POPULATION MEAN USING POPULATION CORRELATION COEFFICIENT

K.B.Panda[#1] and M. Samantaray[#2]

[#]*Department of Statistics, Utkal University, Bhubaneswar, Odisha, India*

1. kunja.st@utkaluniversity.ac.in

2. so.sweet.madhusmita@gmail.com

ABSTRACT

In this paper, a modified product estimator of the finite population mean of variable of interest is suggested using population correlation coefficient between the study variable and the auxiliary variable. The bias and mean squared error of the proposed estimator are obtained. It is found that the estimator is approximately unbiased. It has been shown that the newly suggested estimator based on SRS is more efficient than the simple mean estimator, usual product estimator. To judge the merits of the proposed estimator over other estimators, an empirical study is carried out.

Key Words:

Finite Population Mean, correlation Coefficient, product estimator, Bias and Mean Squared Error.

1: INTRODUCTION

In order to improve the efficiency of the estimators, auxiliary information is used at both selections as well as estimation stage. When the correlation coefficient between the study variate and the auxiliary variate is positive, the ratio method of estimation is used and in case of negative correlation coefficient, product method of estimation is used. Cochran (1940) used auxiliary information at estimation stage and proposed ratio estimator, Murthy (1964) envisaged product estimator and Searl (1964), Sisodia and Dwibedi (1981) utilized co-efficient of variation of auxiliary variable in their respective ratio and product method of estimation. Srivenkataraman (1980) first proposed dual to ratio estimator, Singh and Tailor (2005) and Tailor and Sharma (2009) worked on ratio cum product estimator. Deriving inspiration from the above works completed with the estimator due to Mallick and Tailor (2013) and Panda and Samantaray (2018), we have proposed a new product -cum-dual to product estimator of finite population mean.

Consider a finite population U: $U_1, U_2, \ldots\ldots, U_N$ of N units. Let (y_i, x_i), i = 1, 2...n denote the values of the units included in a sample of size n drawn by simple random sampling without replacement (SRSWOR). In order to have a survey estimate of the population mean $\bar{Y}$ of the study variable Y, assuming the knowledge of the population mean $\bar{X}$ of the auxiliary variable X, Murthy(1964) proposed the classical product estimator defined as

$$\hat{\bar{Y}}_p = \bar{y}\frac{\bar{x}}{\bar{X}} \tag{1.1}$$

where, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are unbiased estimator of population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{n} Y_i$ and $\bar{X} = \frac{1}{N}\sum_{i=1}^{n} X_i$ respectively.

Then, the bias and MSE of $\hat{\bar{Y}}_p$ up to first degree of approximation are obtained as

$$Bias(\hat{\bar{Y}}_p) = \bar{Y} f_1 \rho C_y C_x = \bar{Y} f_1 C_{yx} \tag{1.2}$$

$$MSE(\hat{\bar{Y}}_p) = \bar{Y}^2 f_1 (C_y^2 + C_x^2 + 2C_{yx}) = \bar{Y}^2 f_1 (C_y^2 + C_x^2(1+2k)) \tag{1.3}$$

Utilizing information on correlation co-efficient ρ between the study variable Y and auxiliary variable X Singh and Tailor (2003) the suggested ratio type estimator

$$\hat{\bar{Y}}_r^* = \bar{y}\frac{(\bar{X}+\rho)}{(\bar{x}+\rho)}. \tag{1.4}$$

And in double sampling, the corresponding ratio estimator is expressed as

$$\hat{\bar{Y}}'^{*}_{r} = \bar{y}\frac{(\bar{x}' + \rho)}{(\bar{x} + \rho)}. \quad (1.5)$$

Where, $f_1 = \frac{1}{n} - \frac{1}{N}$,

$$\sigma_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2},$$

$$\sigma_X = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2},$$

$$\sigma_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}),$$

$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ = population correlation coefficient,

$C_{yx} = \rho C_Y C_X$,

$C_X^2 = \frac{\sigma_X^2}{\bar{X}^2}$ And $C_Y^2 = \frac{\sigma_Y^2}{\bar{Y}^2}$ are co-efficient of variation of X and Y respectively.

Panda and Samantaray (2018) proposed a new product type estimator as

$$\hat{\bar{Y}}^{*}_{p} = \bar{y}\frac{(\bar{x} + \rho)}{(\bar{X} + \rho)} \quad (1.6)$$

The bias and MSE of the estimators (up to first degree of approximation) are

$$Bias(\hat{\bar{Y}}^{*}_{p}) = E(\bar{y}^{*}_{p} - \bar{Y}) = \theta.\bar{Y}.f_1.\rho.C_Y.C_X = \theta.\bar{Y}.f_1.C_{YX} \quad (1.7)$$

$$MSE(\hat{\bar{Y}}^{*}_{p}) = E(\hat{\bar{Y}}^{*}_{p} - \bar{Y})^2 = \bar{Y}.f_1[C_Y^2 + \theta C_X^2(\theta + 2\rho.{}^{C_Y}\!/_{C_X}) = \bar{Y}.f_1[C_Y^2 + \theta C_X^2(\theta + 2k)] \quad (1.8)$$

2.1. THE SUGGESTED ESTIMATOR:

We propose a new product-type estimator using correlation coefficient for one auxiliary variable as

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

To derive the expressions for Bias and MSE of the suggested estimator $\hat{\bar{Y}}'_{p}$, we let

$e_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}$, such that $E(e_0)=0$ and $E(e_0^2)= f_1 C_Y^2$ and

$e_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$, such that $E(e_1)=0$ and $E(e_1^2)= f_1 C_X^2$

As our population is infinitely large, so we can drop finite population correction factor (fpc).

Then equation (2.1) becomes $\hat{\bar{Y}}_p^{'} = \bar{Y}(1+e_0)(1+\alpha\theta e_1)$ (2.2)

Assume that $|\theta e_1|<1$ and α is a suitable constant, so that equation (2.2) can be expanded. So, expanding above equation up to second power of $e's$,

$$\text{we have } (\hat{\bar{Y}}''_p - \bar{Y}) = \bar{Y}(1+e_0+\alpha\theta e_1 e_0) \qquad (2.3)$$

Taking expectation on both sides,

$$\text{E } (\hat{\bar{Y}}''_p - \bar{Y}) = f_1 \bar{Y}(2\alpha\theta\rho C_x C_Y) = f_1 \bar{Y}(\alpha.\theta.C_{YX}) = B(\hat{\bar{Y}}_p') \ (4.25)$$

Considering above equation upto first order of approximation, we have

$$\hat{\bar{Y}}''_p - \bar{Y} = (e_0 + \alpha\theta e_1) \qquad (2.4)$$

Squaring and taking expectation on both sides, we have

$$\text{MSE}(\hat{\bar{Y}}''_p) \cong f_1[\sigma_Y^2 + \alpha^2\theta^2 C_X^2 + 2\alpha\theta C_{YX}]$$

$$\cong \bar{Y}^2 f_1[C_Y^2 + \alpha\theta C_X^2(\alpha\theta + 2k)] \qquad (2.5)$$

2.2.1: Efficiency comparison:

Differentiating (2.5) with respect to α and equating it to zero, we have

$$\alpha_{opt} = -\frac{k}{\theta} = -\frac{\bar{X}\sigma_{XY}}{\bar{Y}\theta\sigma_X^2}. \qquad (2.6)$$

Substituting the above value of α in equation (2.5), we arrive at

$$Min(MSE(\hat{\bar{Y}}_p')) = f_1\sigma_Y^2(1-\rho^2), \qquad (2.7)$$

it is equal to the MSE of the linear regression estimator. Thus the newly proposed estimator serves as an alternative to the linear regression estimator when α is chosen optimally.

I. Comparison of Biases

The Bias of the newly proposed product-type estimator is less than that of the customary product estimator iff

$$\left|Bias(\hat{\bar{Y}}''_p)\right| \le \left|Bias(\hat{\bar{Y}}_p)\right|$$

$$\Rightarrow \left|f_1\bar{Y}\alpha\theta C_{YX}\right| \le \left|f_1\bar{Y}C_{YX}\right|$$

$$\Rightarrow |\alpha\theta| \le 1 . \qquad (2.8)$$

II. Comparison of Min (MSE($\hat{\bar{Y}}''_p$)) with MSE($\hat{\bar{Y}}_p$)

Comparing (2.5) with (1.1), we find that the estimator $\hat{\bar{Y}}'_p$ would be more efficient than the estimator $\hat{\bar{Y}}_p$ iff

$$Min(MSE(\hat{\bar{Y}}''_p)) - MSE(\hat{\bar{Y}}_p) = \bar{Y}^2 f_1 C_Y^2(1-\rho^2) - \bar{Y}^2 f_1(C_Y^2 + C_X^2(1+2k)) \le 0$$

$$\Rightarrow \bar{Y}^2 f_1 C_X^2 (\rho + \frac{C_Y}{C_X})^2 \ge 0$$

$$\Rightarrow (\rho + \frac{C_Y}{C_X})^2 \ge 0 \text{, which is always true.}$$

So, $\hat{\bar{Y}}''_p$ is invariably more efficient than $\hat{\bar{Y}}_p$. (2.9)

2.2. SUGGESTED ESTIMATOR IN TWO-PHASE SAMPLING:

When $\bar{X}$ is unknown, We have to consider two-phase sampling or double sampling procedure, where we replace $\bar{X}$ by $\bar{x}'$, i.e., the sample mean based on large preliminary sample of size n' drawn with SRSWOR from population of size N.

$$\hat{\bar{Y}}_p' = \bar{y}\frac{\bar{x}}{\bar{x}'} \qquad (2.10)$$

and $$\hat{\bar{Y}}'^*_p = \bar{y}\frac{(\bar{x}+\rho)}{(\bar{x}'+\rho)} \qquad (2.11)$$

Where, $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$ is unbiased estimator of population mean $\bar{X}$ based on sample size n'.

Thus, the bias and MSE of $\hat{\bar{Y}}'_p$ up to first degree of approximation are obtained as,

$$Bias(\hat{\bar{Y}}'_p) = \bar{Y} f_3 \rho C_Y C_X = \bar{Y} f_3 C_{YX} \ , \qquad (2.12)$$

$$MSE(\hat{\bar{Y}}'_p) = \bar{Y}[f_1 C_Y^2 + C_X^2 f_3 (1 + 2\rho \, C_y / C_X)] = \bar{Y}[f_1 C_Y^2 + C_X^2 f_3 (1+2k)],$$

Here $f_2 = \frac{1}{n'} - \frac{1}{N}$, $f_3 = f_1 - f_2 = \frac{1}{n} - \frac{1}{n'}$, $k = \rho \, C_Y / C_X$ (2.13)

The bias and MSE of $\bar{y}_p^{*'}$ up to first degree of approximation are defined as,

$$Bias(\hat{\bar{Y}}'^{*}_p) = E(\bar{y}'^{*}_p - \bar{Y}) = \theta . \bar{Y} . f_3 . \rho . C_Y . C_X = \theta . \bar{Y} . f_3 . C_{YX} \qquad (2.14)$$

$$MSE(\hat{\bar{Y}}'^{*}_p) = E(\bar{y}'^{*'}_p - \bar{Y})^2 = \bar{Y} . [f_1 C_Y^2 + \theta C_X^2 f_3 (\theta + 2\rho . C_Y / C_X) = \bar{Y}[f_1 C_Y^2 + \theta C_X^2 f_3 (\theta + 2k)]$$

(2.15)

It can be easily be seen that, in two phase sampling , the performance of the propose estimator as measured in terms of bias and mean square error is better than its competing estimator.

The proposed product-type estimator in two phase sampling are as follows:

$$\hat{\bar{Y}}_p^{**} = \bar{y}\left[\frac{\beta(\bar{x}+\rho)+(1-\beta)(\bar{x}'+\rho)}{(\bar{x}'+\rho)}\right], \qquad (2.16)$$

where $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$ is the unbiased estimator of $\bar{X}$ based on sample size n' and β is a suitable constant .

where $C'_x = \frac{\sigma'_x}{\bar{x}'}$, $\rho' = \frac{\sigma_{Y\bar{x}'}}{\sigma_{\bar{x}'}\sigma_Y}$ and $\theta' = \frac{\bar{x}'}{\bar{x}'+\rho'}$

Let $e'_1 = \frac{\bar{x} - \bar{x}'}{\bar{x}'}$, such that $E(e'_1) = 0$ and $E(e'^2_1) = f_3 C'^2_{x'}$

The Bias and MSE of $\hat{\bar{Y}}_p^{**}$ up to first degree of approximation are as follows,

$$Bias(\hat{\bar{Y}}_p^{**}) = \bar{Y} f_3 \beta \theta' C'_{Yx'} \ , \qquad (2.17)$$

$$MSE(\hat{\bar{Y}}_p^{**}) = \bar{Y}^2[f_1 C_Y^2 + \beta\theta' C'^2_{x'} f_3 (\beta\theta' + 2\rho' \, C_Y / C'_{x'})] = \bar{Y}^2[f_1 C_Y^2 + \beta\theta' C'^2_{x'} f_3 (\beta\theta' + 2k')] \qquad (2.18)$$

2.2.2: Efficiency comparison:

The optimum value of β can be arrived at by differentiating (2.18) and equating it to zero. Thus

$$\beta_{opt} = -\frac{k'}{\theta'} = -\frac{\bar{x}'\sigma_{x'Y}}{\bar{Y}\theta'\sigma^2_{x'}}. \qquad (2.19)$$

Substituting the above value of β_{opt} in equation (2.18), we arrive at

$$Min(MSE(\hat{\bar{Y}}^{**}_p)) = \bar{Y}^2 C^2_Y (f_1 - f_3\rho'^2). \qquad (2.20)$$

I. Comparison of Biases

The Bias of the newly proposed product-type estimator in second stage sampling is less than that of the customary product estimator in second stage sampling iff

$$\left|Bias(\hat{\bar{Y}}^{**}_p)\right| \le \left|Bias(\hat{\bar{Y}}^{*}_p)\right|$$

$$\Rightarrow \left|f_3\bar{Y}\beta\theta' C'_{Yx'}\right| \le \left|f_3\bar{Y}C'_{Yx'}\right|$$

$$\Rightarrow \left|\beta\theta'\right| \le 1. \qquad (2.21)$$

II. Comparison of Min (MSE($\hat{\bar{Y}}^{**}_p$)) with MSE($\hat{\bar{Y}}'_p$)

Comparing (2.20) with (2.15), we find that the estimator $\hat{\bar{Y}}^{**}_p$ would be more efficient than the estimator $\hat{\bar{Y}}'_p$ iff

$$Min(MSE(\hat{\bar{Y}}^{**}_p)) - MSE(\hat{\bar{Y}}'_p) = \bar{Y}^2 C^2_Y (f_1 - f_3\rho'^2) - \bar{\bar{Y}}^2 (f_1 C^2_Y + f_3 C'^2_{x'}(1+2k')) \le 0$$

$$\Rightarrow \bar{Y}^2 f_3 C'^2_{x'} (\rho' + \frac{C_Y}{C'_x})^2 \ge 0$$

$$\Rightarrow (\rho' + \frac{C_Y}{C'_{x'}})^2 \ge 0$$, which is always true.

Thus, $\hat{\bar{Y}}^{**}_p$ is found to be more efficient than $\hat{\bar{Y}}'_p$ unconditionally. (2.22)

3. EMPIRICAL STUDY:

To obtain performance of suggested estimator $\hat{\bar{Y}}''_p$ over its competitors, a natural population data set is being considered.

We refer to Example8.1 (Highway data) given in Weisberg (1980, p.179), wherein the sample quantities have been taken as the corresponding population quantities and are furnished hereunder:

Y: RATE=1973 accident rate per million vehicle miles,

X: LEN= Length of the segment in miles,

Z: ADT= Average daily traffic count in thousands (estimated),

n=30, $\bar{Y}$ =3.933, $\bar{X}$ =12.88, $\bar{Z}$ =19.62

σ_Y^2=3.944, σ_X^2 =57.91, σ_Z^2=34.64, σ_Y =1.986, σ_X =7.610, σ_Z=18.61, ρ =(-0.47),

Then using above data, we have α_{opt} =0.38704.

The Bias, MSE and PRE of the above estimators have been computed and presented in the following table:

Table 1: BIAS, MSE and PRE of the Competing Estimators

SL. No.	ESTIMATORS	$\lvert BIAS \rvert$	MSE	PRE
1	$\hat{\bar{Y}}$	0.000	0.7694	100.00
2	$\hat{\bar{Y}}_p$	0.0003	0.2493	308.64
3	$\hat{\bar{Y}}''_p$	0.000042	0.2175	353.76

From the above table, it is found that gain in efficiency of the proposed estimator $\hat{\bar{Y}}''_p$ with respect to $\hat{\bar{Y}}$ and $\hat{\bar{Y}}_p$ are 308.64 and 353.76, implying thereby that there is a substantial increase in gain in efficiency of proposed estimator over its competitors with less MSE.

CONCLUSION:

The proposed modified product type estimator is more efficient than simple mean; it fares better than the classical product estimator under feasible conditions. Empirical study based on real data set is in agreement with the theoretical findings.

REFERENCES:

[1] Cochran W.G., 1977: Sampling Techniques. Third U.S. Edition. Wiley Eastern Limited. A.; 325

[2] Cochran W. G., Sampling Theory when sampling Units are of Unequal sizes. Journal American Statist. Association. 1942. 37; 191-212

[3] Das A. K., 1988: Contribution to the Theory of Sampling Strategies based on Auxiliary Information. Ph. D. Thesis, BCKY, West Bengal, India.

[4] Housila P. S., Tailor R. and Tailor R., 2012: Estimation of Finite Population Mean In Two Phase Sampling with Known Coefficient of Variation of Auxiliary Character. Statistica. 72(1) ;1-22

[5] Kalidar C. and Cingi H., 2004: Ratio Estimators in Simple Random sampling. Applied Mathematics and Computation. 151; 893-902

[6] Malik K. A. and Tailor R., 2013: Ratio Type Estimator Of Population Mean in double sampling. International Jour. Of Advanced Math. and Statist. Volume 1.Issue 1; 34-39

[7] Murthy M. N., 1964: Product Method Of Estimation. Sankhya. A. 26; 294-307.

[8] Panda, K. B. and Samantaray, M. 2018: "A New Product estimator of population mean and its efficiency" , Int. J. of Mathematical Trends and Technology, Vol-53, Part-IV, 37-42.

[9] Searls D.T., 1964: The Utilization of Known Coefficient of Variation in the estimation procedure. Journal of American Statistical Association. 59; 1125-1126

[10] Singh H. P. and Tailor R. , 2005: Estimation of Population Mean Using Known Correlation Coefficient between Auxiliary Characters. Statistica. Anno LXV. 4; 407-418

[11]Sisodia B. V. S. and Dwibedi V. K., 1981: A Modified Ratio Estimator Using Co-efficient of Variation Of Auxiliary variable. Journal. Ind. Soc. Agril. Statist. 33(2); 13-18

[12]Sodipo A. A. and Obisesan K. O., 2007: Estimation of The Population Mean Using Difference Cum Ratio Estimator with Full Response on the Auxiliary Character. Research Journal Of Applied Sciences. 2(6); 769-772

[13]Srivenkataramana T., 1980: A Dual Of Ratio Estimator in sample surveys. Biometrika.67(1). 199-204

[14]Tailor R. and Sharma B. K., 2009: A Modified Ratio-Cum- Product Estimator Of Finite Population Mean Using Known Co-efficient Of Variation and Co-efficient Of Kurtosis. Statistics in Transition-New Series. 10(1). 15-24

[15] Weisberg, S. 1980: "Applied linear regression analysis", John Wiley and Sons, Inc, New York.