

Original Research Article

<https://doi.org/10.20546/ijcmas.2021.1002.096>

Utilizing Sample Size Information for Improved Estimation of Population Mean in Agriculture Surveys

D. K. Yadav¹, Shailendra Kumar^{2*} and Tanuj Misra²

¹National Institute of Health & Family Welfare, New Delhi, India

²Rani Lakshmi Bai Central Agricultural University, Jhansi, U.P., India

*Corresponding author

ABSTRACT

Keywords

Estimation,
Bias, MSE

Article Info

Accepted:
07 January 2021
Available Online:
10 February 2021

Sample size is a very important feature for estimation of population parameters in various socio-economic and agricultural surveys. Precision of population estimates depends on adequate sample size. In the present paper, we are developing an improved estimation technique for estimating population mean using sample size as auxiliary information. The asymptotic properties of proposed estimator like bias and mean square error (MSE) have been obtained up to first order of approximation. An efficiency comparison is also made between proposed estimator and other competing estimators. To justify theoretical results an empirical study is also carried out.

Introduction

The reliable estimates play a vital role in effective policy making of various agriculture related programmes. Successful implementation of agriculture schemes depends on reliable crop estimates. These estimates are obtained from various agriculture related surveys like CCS (Crop Cutting Surveys). Use of auxiliary information is a very common practice for increasing the efficiency of the estimation techniques in these surveys. Suppose we are conducting a agriculture survey regarding the production of any particular crop. Here population total is an important parameter which is used for assessment of agriculture

production. In such cases crop production and total area under cultivation are positively correlated variables so in sampling theory ratio estimators will give more reliable estimates in this situation. Population mean is one of the very important measures of central tendency used in almost all fields of society including field of agriculture. Thus the estimation of population mean is of great significance in agriculture related areas. In the present manuscript, a modified ratio type estimator of population mean of the study variable using information on size of the sample has been proposed and its large sample properties have been studied up to the first order of approximation.

Notations & Terminology

Let the finite population under consideration consist of N distinct and identifiable units and let (x_i, y_i) $i = 1, 2, \dots,$ be a bivariate sample of size n taken from (X, Y) using a SRSWOR scheme. Let \bar{X} and \bar{Y} respectively be the population means of the auxiliary and the study variables and \bar{x} and \bar{y} be the corresponding sample means. It is well known and has been seen in practice that in simple random sampling scheme, sample means \bar{x} and \bar{y} are unbiased estimators of population means of \bar{X} and \bar{Y} respectively.

Review of existing estimators

The usual and the most suitable estimator for estimating population mean \bar{Y} is sample mean \bar{y} , given by

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \tag{1}$$

It is unbiased for population mean and its variance upto the first order of approximation is given by

$$V(\bar{y}) = \gamma S_y^2 = \gamma \bar{Y}^2 C_y^2$$

Where, $C_y = \frac{S_y}{\bar{Y}}$, $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$

$$\gamma = \frac{N-n}{nN}$$

Cochran (1940) used the positively correlated auxiliary variable with study variable and proposed the following usual ratio estimator of population mean as,

$$\hat{\bar{Y}}_R = \bar{y} \frac{\bar{X}}{\bar{x}}$$

The above estimator is biased estimator of population mean and its bias and MSE, upto the first order of approximation respectively are,

$$B(\hat{\bar{Y}}_R) = \gamma \bar{Y} [C_x^2 - \rho_{yx} C_y C_x]$$

$$MSE(\hat{\bar{Y}}_R) = \gamma \bar{Y}^2 [C_y^2 + C_x^2 - 2\rho_{yx} C_y C_x] \tag{4}$$

Where, $C_x = \frac{S_x}{\bar{X}}$, $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$, $\rho_{yx} = \frac{Cov(x,y)}{S_x S_y}$,

$$Cov(x,y) = \frac{1}{(N-1)} \sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}), \quad \gamma = \frac{N-n}{nN}$$

In literature various modified estimators of population mean of study variable using auxiliary variables have been given by various authors. For detailed study of the modified ratio type estimators, latest references can be made of Kadilar and Cingi (2004, 2009), Singh (2003), Subramani (2013), Tailor and Sharma (2009), Yadav and Adewara (2013), Yadav *et al.*, (2016a, 2016b, 2016c, 2016d), Yadav *et al.*, (2017a, 2017b, 2017c), Yadav and Gupta (2017, 2018), Yadav *et al.*, (2018a, 2018b, 2018c), Yadav *et al.*, (2019) and Jaroengeratikun and Lawson (2019)

(2)

Following table-1 represents some modified estimators, their constants, biases and mean squared errors.

Proposed estimator

Motivated by Jaroengeratikun and Lawson (2019) estimator of population mean, we propose the following ratio type estimator of the population mean using information on the on size of the sample as,

$$(3) \hat{t}_R = \bar{y} \left[\frac{a\bar{X} + c}{a\bar{x} + c} \right]$$

Where c is any known constant. In our research paper we will take $c = n$ (sample size).

Thus proposed estimator will be,

$$\tau_R = \bar{y} \left[\frac{a\bar{X} + n}{a\bar{X} + n} \right] \quad (5)$$

To study the large sample properties of the estimator, we define

$$\bar{y} = \bar{Y}(1 + \varepsilon_0) \text{ and } \bar{x} = \bar{X}(1 + \varepsilon_1) \text{ such that } E(\varepsilon_i) = 0 \text{ (i = 0,1)} \text{ and } E(\varepsilon_0^2) = \gamma C_y^2, \\ E(\varepsilon_1^2) = \gamma C_x^2, E(\varepsilon_0 \varepsilon_1) = \gamma C_x C_y.$$

Expressing (5) in terms of ε_i 's, we have

$$\tau_R = \bar{Y} (1 + \varepsilon_0) \left[\frac{a\bar{X} + n}{a\bar{X} (1 + \varepsilon_1) + n} \right] \\ = \bar{Y} (1 + \varepsilon_0) \left[\frac{1}{1 + \left(\frac{a\bar{X}}{a\bar{X} + n} \right) \varepsilon_1} \right] \\ = \bar{Y}(1 + \varepsilon_0) \left[\frac{1}{(1 + \delta\varepsilon_1)} \right], \text{ where } \delta = \left(\frac{a\bar{X}}{a\bar{X} + n} \right) \\ = \bar{Y}(1 + \varepsilon_0)(1 + \delta\varepsilon_1)^{-1}$$

We assume that, $I_{\varepsilon_1} < 1$, so that $(1 + \delta\varepsilon_1)^{-1}$ may be expanded

$$= \bar{Y}(1 + \varepsilon_0)(1 - \delta\varepsilon_1 + \delta^2\varepsilon_1^2) \\ = \bar{Y} + \bar{Y}[\varepsilon_0 - \delta\varepsilon_0\varepsilon_1]$$

Retaining the terms upto the first order of approximations, we have

$$\tau_R - \bar{Y} = \bar{Y}[\varepsilon_0 - \delta\varepsilon_0\varepsilon_1] \quad (6)$$

Taking expectation both sides and putting the values of different expectations, we get the bias of proposed estimator upto the first order of approximation

$$E(\tau_R - \bar{Y}) = \bar{Y}[E(\varepsilon_0) - \delta E(\varepsilon_0\varepsilon_1)]$$

$$\text{Bias}(\tau_R) = -\gamma\bar{Y} [\delta C_x C_y]$$

Now, squaring (6) and taking expectations both sides we get MSE of τ_R upto the first order of approximation

$$E(\tau_R - \bar{Y})^2 = \bar{Y}^2 E[\varepsilon_0 - \delta\varepsilon_0\varepsilon_1]^2 \\ = \bar{Y}^2 E[\varepsilon_0^2 + \delta^2\varepsilon_0\varepsilon_1^2 - 2\delta\varepsilon_0^2\varepsilon_1]$$

Retaining the terms upto the first order of approximations, we have

$$E(\tau_R - \bar{Y})^2 = \bar{Y}^2 E(\varepsilon_0^2)$$

$$\text{MSE}(\tau_R) = \gamma\bar{Y}^2 C_y^2$$

Theoretical efficiency comparison

Proposed estimator τ_R will be more efficient than usual ratio estimator $\hat{\bar{Y}}_R$, if

$$\text{MSE}(\hat{\bar{Y}}_R) - \text{MSE}(\tau_R) > 0$$

gives $\rho < \frac{1}{2} \left(\frac{C_x}{C_y} \right)$

Proposed estimator τ_R will be more efficient than the estimator $\hat{\bar{Y}}_1$ proposed by Sisodia and Dwivedi (1981), if

$$\text{MSE}(\hat{\bar{Y}}_1) - \text{MSE}(\tau_R) > 0$$

gives $\rho < \frac{1}{2} \left(\frac{\delta_x C_x}{C_y} \right)$

Proposed estimator τ_R will be more efficient than the estimator $\hat{\bar{Y}}_2$ proposed by Upadhyaya and Singh (1999), if

$$MSE(\hat{Y}_2) - MSE(\tau_R) > 0$$

Gives
$$\rho < \frac{1}{2} \left(\frac{\delta_2 C_x}{C_y} \right)$$

- I. Proposed estimator τ_R will be more efficient than the estimator \hat{Y}_3 proposed by Singh and Tailor(2003), if

$$MSE(\hat{Y}_3) - MSE(\tau_R) > 0$$

Gives
$$\rho < \frac{1}{2} \left(\frac{\delta_3 C_x}{C_y} \right)$$

- II. Proposed estimator τ_R will be more efficient than the estimator \hat{Y}_4 proposed by Singh *et al.*, (2004), if

$$MSE(\hat{Y}_4) - MSE(\tau_R) > 0$$

Gives
$$\rho < \frac{1}{2} \left(\frac{\delta_4 C_x}{C_y} \right)$$

- III. Proposed estimator τ_R will be more efficient than the estimator \hat{Y}_5 proposed by Yan and Tian (2010), if

Gives
$$\rho < \frac{1}{2} \left(\frac{\delta_5 C_x}{C_y} \right)$$

Empirical study

To judge the performances of the proposed estimator and the existing estimators of population mean using auxiliary variable, we have considered four natural populations from two sources. First two populations, population-1 and population-2 are from Murthy (1967) and rest two populations, population-3 and population-4 are from Mukhopadhyay (2009).

Murthy (1967)

Population 1: Y = Output for 80 factories in a region and X = Number of workers

$$X = 80, n = 20, \bar{Y} = 51.8264, \bar{X} = 11.2646$$

$$\rho = .9413, C_y = 0.3542, C_x = .07507$$

$$\beta_1 = 1.0500, \beta_2 = -0.0634, M_d = 7.5750$$

Population 2: Y = Output for 80 factories in a region and X = Fixed Capital

$$N = 80, n = 20, \bar{Y} = 51.8264, \bar{X} = 2.8153$$

$$\rho = 0.9150, C_y = .3542, C_x = 0.9485$$

$$\beta_1 = 1.3006, \beta_2 = 0.6977, M_d = 1.4800$$

Mukhopadhyay (2009)

Population 3: Y = Output for 40 factories in a region and X = Number of workers

$$N = 40, n = 8, \bar{Y} = 50.7858, \bar{X} = 2.3033$$

$$\rho = 0.8006, C_y = 0.3295, C_x = 0.8406$$

$$\beta_1 = 0.9740, \beta_2 = -0.5344, M_d = 1.250$$

Population 4: Y = Output for 40 factories in a region and X = Fixed Capital

$$N = 40, n = 8, \bar{Y} = 50.7858, \bar{X} = 9.4543$$

$$\rho = 0.8349, C_y = 0.3295, C_x = 0.6756$$

$$\beta_1 = 0.8799, \beta_2 = -0.4622, M_d = 7.0700$$

Following Table-2 and Table-3 represents the biases, mean squared errors of proposed and existing estimators of population mean.

Table.1 Various estimators, their constants, biases and mean squared errors

S.No	Estimators	Constants	Bias	MSE
1	$\bar{y}_1 = \bar{y} \left(\frac{\bar{X} + C_x}{\bar{x} + C_x} \right)$ Sisodia and Dwivedi [1981]	$\delta_1 = \left(\frac{\bar{X}}{\bar{x} + C_x} \right)$	$\gamma \bar{Y} (\delta_1^2 C_x^2 - 2\delta_1 \rho C_x C_y)$	$\gamma \bar{Y}^2 (C_y^2 + \delta_1^2 C_x^2 - 2\delta_1 \rho C_x C_y)$
2	$\bar{y}_2 = \bar{y} \left(\frac{\bar{X} C_x + \beta_2}{\bar{x} C_x + \beta_2} \right)$ Upadhyaya & Singh (1999)	$\delta_2 = \left(\frac{\bar{X} C_x}{\bar{x} C_x + \beta_2} \right)$	$\gamma \bar{Y} (\delta_2^2 C_x^2 - 2\delta_2 \rho C_x C_y)$	$\gamma \bar{Y}^2 (C_y^2 + \delta_2^2 C_x^2 - 2\delta_2 \rho C_x C_y)$
3	$\bar{y}_3 = \bar{y} \left(\frac{\bar{X} + \rho}{\bar{x} + \rho} \right)$ Singh & Tailor (2003)	$\delta_3 = \left(\frac{\bar{X}}{\bar{x} + \rho} \right)$	$\gamma \bar{Y} (\delta_3^2 C_x^2 - 2\delta_3 \rho C_x C_y)$	$\gamma \bar{Y}^2 (C_y^2 + \delta_3^2 C_x^2 - 2\delta_3 \rho C_x C_y)$
4	$\bar{y}_4 = \bar{y} \left(\frac{\bar{X} + \beta_2}{\bar{x} + \beta_2} \right)$ Singh et al. (2004)	$\delta_4 = \left(\frac{\bar{X}}{\bar{x} + \beta_2} \right)$	$\gamma \bar{Y} (\delta_4^2 C_x^2 - 2\delta_4 \rho C_x C_y)$	$\gamma \bar{Y}^2 (C_y^2 + \delta_4^2 C_x^2 - 2\delta_4 \rho C_x C_y)$
5	$\bar{y}_5 = \bar{y} \left(\frac{\bar{X} + \beta_1}{\bar{x} + \beta_1} \right)$ Yan and Tian (2010)	$\delta_5 = \left(\frac{\bar{X}}{\bar{x} + \beta_1} \right)$	$\gamma \bar{Y} (\delta_5^2 C_x^2 - 2\delta_5 \rho C_x C_y)$	$\gamma \bar{Y}^2 (C_y^2 + \delta_5^2 C_x^2 - 2\delta_5 \rho C_x C_y)$

Table.2 Biases of the existing and proposed modified ratio estimators for four natural populations

Estimator	Population I	Population II	Population III	Population IV
\bar{y}_R	0.60	1.15	2.46	1.37
\bar{y}_1	0.05	0.08	0.27	0.25
\bar{y}_2	0.11	0.15	3.73	0.65
\bar{y}_3	0.03	0.09	0.30	0.22
\bar{y}_4	0.11	0.16	3.15	0.57
\bar{y}_5	0.11	0.00	0.18	0.21
τ_R (Proposed)	0.18	0.08	0.31	0.61

Table.3 Variance/mean squared errors of the existing and proposed modified ratio estimators for four natural populations

Estimator	Population I	Population II	Population III	Population IV
\bar{y}_R	18.97	41.32	95.86	49.85
\bar{y}_1	15.25	17.19	42.01	41.06
\bar{y}_2	18.51	20.67	217.70	61.45
\bar{y}_3	14.45	17.69	43.47	39.30
\bar{y}_4	18.62	21.37	188.05	57.33
\bar{y}_5	18.70	12.84	37.62	38.82
τ_R (Proposed)	12.62	12.63	27.98	28.00

Results and Discussion

In the present manuscript we have proposed a generalized ratio type estimator of the study variable by making use of information on the size of the sample. Table-2 and Table-3 represent the biases and the mean squared errors of the proposed and the existing estimators. From both the tables, we can see that the proposed estimator has minimum bias and mean squared error in all four natural populations. Thus proposed estimator is recommended to survey practitioners for its use in various agriculture surveys.

References

- Cochran, W. G. The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *The Journal of Agric. Science*, 1940. 30: 262-275.
- Gupta, R.K. and Yadav, S.K. Improved Estimation of Population Mean Using Information on Size of the Sample, *American Journal of Mathematics and Statistics*, 2018, 8(2):27-35
- Jaroengeratikun, U. and Lawson, N. A combined family of ratio estimators for population mean using an auxiliary variable in simple random sampling, *Journal of Mathematics and Fundamental Sciences*. 2019, Vol. 51 (1): 1–12.
- Jerajuddin, M., and Kishun, J. Modified Ratio Estimators for Population Mean Using Size of the Sample, *Selected From Population*, *IJSRSET*. 2016, 2(2): 10-16.
- Kadilar, C., and Cingi, H. *Advances in sampling theory - Ratio method of estimation*, Bentham Science Publishers, 2009.
- Kadilar, C., and Cingi, H. Ratio estimators in simple random sampling, *Applied Mathematics and Computation*, 2004. 151: 893–902
- Koyuncu, N., Kadilar, C. Efficient estimators for the population mean, *Hacettepe Journal of Mathematics and Statistics*, 2009. 38(2): 217–225.
- Mukhopadhyay, P. *Theory and methods of survey sampling*, PHI Learning, 2nd edition, New Delhi, 2009.
- Murthy, M.N. *Sampling Theory and Methods*; Statistical Publishing Society, Calcutta. 1967.
- Singh, D., and Chaudhary, F. S.. *Theory and analysis of sample survey designs*, New Age International Publisher, New Delhi, 1986.
- Singh, H. P. and Tailor, R. Use of known correlation coefficient in estimating the finite population means, *Statistics in Transition*. 2003. 6 (4): 555–560.
- Sisodia, B.V.S. and Dwivedi, V.K. A Modified Ratio Estimator using Coefficient of Variation of Auxiliary Variable; *Jour. of Indian. Soc. of Agri. Stat.*, 1981, 33(1): 13–18.
- Upadhyaya, L.N. and Singh, H.P. Use of transformed auxiliary variable in estimating the finite population means, *Biometrical Journal*. 1999. 41 (5): 627–636.
- Yadav, D. K., Kumar, R., Misra, S., Yadav, S.K. An Improved Estimator of Population Mean using Information on Median of the Study Variable, *International Journal of Mathematics Trends and Technology(IJMTT)*. 2017b, 46(2): 118-124.
- Yadav, D.K., Devi, M., Yadav, S.K. Estimation of Finite Population Mean using Known Coefficient of Variation in the Simultaneous Presence of Non - Response and Measurement Errors under Double Sampling Scheme, *Journal of Reliability and Statistical Studies (JRSS)*, 2018c. 11 (1): 51-66
- Yadav, D.K., Dipika, Misra, S. An Efficient

- Estimator for Estimating Finite Population Mean Using Known Median of the Study Variable. *International Journal of Engineering Sciences and Research Technology(IJESRT)*. 2017c, 6(6): 503-509
- Yadav, D.K., Sharma, D.K. Estimation of finite population mean using linear combination of Tri-Mean and Quartile Average of auxiliary variable under predictive modelling approach. *Review of Business and Technology Research*. 2019. 16(1): 16-21.
- Yadav, D.K., Shukla, A.K., Tomer, S., Kumar, B. Predictive Estimation of Finite Population Mean using Coefficient of Kurtosis and Median of an Auxiliary Variable under Simple Random Sampling Scheme, *International Jour. of Mathematical Achive (IJMAA)*. 2018a, 9 (5): 1-7.
- Yadav, D.K., Singh, L., Yadav, S.K. New Efficient Class of Estimators for the Population Variance, *International Journal of Agricultural and Statistical Sciences*, 2018b, 14: 463-468.
- Yadav, S. K. and Adewara, A.A. On Improved Estimation of Population Mean using Qualitative Auxiliary Information, *Mathematical Theory and Modeling*. 2013. 3(11),: 42-50.
- Yadav, S.K., Yadav, O.P, Yadav, D.K. Dexterous Estimation of population Mean in survey sampling under non-response error. *International Journal of Mathematical, Engineering and Management Sciences*. 2019, 4(6): 1307-1324
- Yan, Z. and Tian, B. Ratio Method to the Mean Estimation using Coefficient of Skewness of Auxiliary Variable; *ICICA, Part-II, CCIS 106*. 2010:103 – 110.

How to cite this article:

Yadav, D. K., Shailendra Kumar and Tanuj Misra. 2021. Utilizing Sample Size Information for Improved Estimation of Population Mean in Agriculture Surveys. *Int.J.Curr.Microbiol.App.Sci*. 10(02): 809-815. doi: <https://doi.org/10.20546/ijemas.2021.1002.096>