The Pharma Innovation Journal 2023; 12(4): 1126-1132

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(4): 1126-1132

www.thepharmajournal.com
Received: 02-02-2023
Accepted: 18-03-2023

Manish Kumar
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya, Uttar Pradesh, India

Ashish Tiwari
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya, Uttar Pradesh, India

Sarvesh Kumar Dubey
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya, Uttar Pradesh, India

Praveen Singh
Department of Agricultural Meteorology, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya, Uttar Pradesh, India

Shiv Kumar Rana
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya, Uttar Pradesh, India

Corresponding Author:
Ashish Tiwari
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture & Technology, Ayodhya, Uttar Pradesh, India

A composite class of estimators for a finite population mean in stratified random sampling

Manish Kumar, Ashish Tiwari, Sarvesh Kumar Dubey, Praveen Singh and Shiv Kumar Rana

Abstract
The present paper deals with the formulation of a composite class of estimators for the population mean. The necessary and sufficient conditions (NASCs) for the superiority of the proposed class T as compared to the preliminary estimators are obtained on using the MSE criterion. An empirical analysis is conducted on using some real population data sets in order to assess the relative performances of the proposed class as compared to the preliminary estimators. The theoretical and empirical findings reveal the proposed class of the estimators is dominant over the well-known preliminary estimators.

Keywords: Study variable, auxiliary variable, stratified random sampling, mean square error, percent relative efficiency

1. Introduction
Increasing the sample size will boost an estimator's effectiveness. A sample drawn from a large population at once is rarely representative in the sense that the sample units might be distributed unevenly across the population, hence doing so is not advisable. The effectiveness can be improved by dividing the population into sub-groups (known as strata) and then taking samples from each stratum separately. The procedure of selecting samples in this manner is regarded as stratified random sampling.
The theory of estimation of mean of a target variable (also regarded as study variable) is of huge importance in various diversified fields of study, for instance agriculture, demographic studies, actuarial sciences, and much more. In various practical situations, the population under study is heterogeneous, and in that case, we adopt stratified random sampling to obtain precise estimators for the population mean of the study variable. Some notable contributions towards the development of estimators for the mean in stratified random sampling have been made by (Bahl and Tuteja, 1991) [1], (Upadhyaya and Singh, 1999) [2], (Kadilar and Cingi, 2003) [3], (Shabbir and Gupta, 2006) [4], (Singh *et al*., 2009) [5], (Vishwakarma and Kumar, 2016) [6], and (Kumar and Vishwakarma, 2020) [7]. Some recent works related to the estimation of mean in survey sampling have been made by (Kumar and Tiwari, 2021) [8], (Kumar and Tiwari, 2022) [9], and (Tiwari *et al*., 2023) [10].

2. Some Preliminary Estimators of the Population Mean
Consider a finite population $U=\{U_1,U_2,.....,U_N\}$ consisting of N units, and the population is being sub-divided into L distinct strata with h^{th} stratum consisting of N_h units, $(h=1,2,.....,L)$, such that $\sum_{h=1}^{L} N_h = N.$ Also, let n_h be the size of the sample drawn from the h^{th} stratum by utilizing simple random sampling without replacement (SRSWOR) scheme such that $\sum_{h=1}^{L} n_h = n.$ Let (Y, X) be the study and auxiliary variables, respectively, taking the values (y_{hi}, x_{hi}) on the i^{th} unit of the h^{th} stratum, where $i=1,2,.....,N_h$.
Moreover, the population means of the variables (Y, X) in the h^{th} stratum are $\left(\bar{Y}_h=\sum_{i=1}^{N_h} y_{hi}/N_h, \bar{X}_h=\sum_{i=1}^{N_h} x_{hi}/N_h\right)$, and the corresponding sample means in the h^{th} stratum are $\left(\bar{y}_h=\sum_{i=1}^{n_h} y_{hi}/n_h, \bar{x}_h=\sum_{i=1}^{n_h} x_{hi}/n_h\right)$. Furthermore, the sample means of the variables (Y, X) in stratified random

sampling are $\left(\bar{y}_{st}=\sum_{h=1}^{L}W_h\bar{y}_h,\ \bar{x}_{st}=\sum_{h=1}^{L}W_h\bar{x}_h\right)$, where $W_h=N_h/N$ denotes the stratum weight. Also, $\bar{y}_{st}$ and $\bar{x}_{st}$ are the unbiased estimators of the population means $\bar{Y}=\sum_{h=1}^{L}W_h\bar{Y}_h$ and $\bar{X}=\sum_{h=1}^{L}W_h\bar{X}_h$, respectively.

The separate ratio estimator for population mean $\bar{Y}$ is defined by

$$\bar{y}_{RS}=\sum_{h=1}^{L}W_h\bar{y}_h\left(\frac{\bar{X}_h}{\bar{x}_h}\right) \tag{1}$$

Also, the separate regression estimator for $\bar{Y}$ is defined by

$$\bar{y}_{lrs}=\sum_{h=1}^{L}W_h\left[\bar{y}_h+b_h\left(\bar{X}_h-\bar{x}_h\right)\right] \tag{2}$$

Where $b_h=s_{yxh}/s_{xh}^2$ denotes the sample regression coefficient of *Y* on *X* in the h^{th} stratum. Also, $s_{xh}^2=\frac{1}{n_h-1}\sum_{i=1}^{n_h}\left(x_{hi}-\bar{x}_h\right)^2$ and $s_{yxh}=\frac{1}{n_h-1}\sum_{i=1}^{n_h}\left(y_{hi}-\bar{y}_h\right)\left(x_{hi}-\bar{x}_h\right)$.

The combined ratio estimator for population mean $\bar{Y}$ is defined by

$$\bar{y}_{RC}=\bar{y}_{st}\left(\frac{\bar{X}}{\bar{x}_{st}}\right) \tag{3}$$

The exponential version of (Bahl and Tuteja, 1991) [1] estimator for population mean $\bar{Y}$ in stratified random sampling is defined by

$$\bar{y}_{exp}=\bar{y}_{st}\exp\left(\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right) \tag{4}$$

The combined regression estimator for population mean $\bar{Y}$ is defined by

$$\bar{y}_{lrc}=\bar{y}_{st}+b\left(\bar{X}-\bar{x}_{st}\right) \tag{5}$$

Where $b=\sum_{h=1}^{L}W_h^2\lambda_h s_{yxh}/\sum_{h=1}^{L}W_h^2\lambda_h s_{xh}^2$, $\lambda_h=\left\{(1/n_h)-(1/N_h)\right\}$.

To the first order of approximation, the variance of stratified sample mean ($\bar{y}_{st}$) and the mean square errors (MSEs) of various above mentioned preliminary estimators are given by

$$\mathrm{Var}\left(\bar{y}_{st}\right)=\sum_{h=1}^{L}W_h^2\lambda_h S_{yh}^2 \tag{6}$$

$$\mathrm{MSE}\left(\bar{y}_{RS}\right)=\sum_{h=1}^{L}W_h^2\lambda_h\left(S_{yh}^2+R_h^2S_{xh}^2-2R_hS_{yxh}\right) \tag{7}$$

$$\mathrm{MSE}\left(\bar{y}_{lrs}\right)=\sum_{h=1}^{L}W_h^2\lambda_h S_{yh}^2\left(1-\rho_{yxh}^2\right) \tag{8}$$

$$\text{MSE}\left(\bar{y}_{\text{RC}}\right)=\sum_{h=1}^{L}W_h^2\lambda_h\left(S_{yh}^2+R^2S_{xh}^2-2RS_{yxh}\right) \tag{9}$$

$$\text{MSE}\left(\bar{y}_{\exp}\right)=\sum_{h=1}^{L}W_h^2\lambda_h\left(S_{yh}^2+\frac{R^2}{4}S_{xh}^2-RS_{yxh}\right) \tag{10}$$

$$\text{MSE}\left(\bar{y}_{\text{lrc}}\right)=\sum_{h=1}^{L}W_h^2\lambda_hS_{yh}^2\left(1-\rho_{yx}^2\right) \tag{11}$$

Where

$$R_h=\bar{Y}_h/\bar{X}_h,\ R=\bar{Y}/\bar{X},\ \rho_{yx}=\frac{\sum_{h=1}^{L}W_h^2\lambda_hS_{yxh}}{\sqrt{\left(\sum_{h=1}^{L}W_h^2\lambda_hS_{yh}^2\right)\left(\sum_{h=1}^{L}W_h^2\lambda_hS_{xh}^2\right)}},\ \rho_{yxh}=\frac{S_{yxh}}{S_{yh}S_{xh}},$$

$$S_{yh}^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)^2,\ S_{xh}^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(x_{hi}-\bar{X}_h\right)^2,$$

$$S_{yxh}=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)\left(x_{hi}-\bar{X}_h\right).$$

3. Proposed Class of Estimators

We propose the following composite class of estimators for the population mean ($\bar{Y}$) of the study variable (Y):

$$T=\bar{y}_{st}\left[\alpha\left(\frac{\bar{X}}{\bar{x}_{\text{st}}}\right)+\left(1-\alpha\right)\exp\left(\frac{\bar{X}-\bar{x}_{\text{st}}}{\bar{X}+\bar{x}_{\text{st}}}\right)\right] \tag{12}$$

Where α is a scalar quantity. The optimum value of α is obtained on minimizing the MSE of the proposed class T.

Remark 3.1: For $\alpha=1,$ the proposed class T reduces to the combined ratio estimator ($\bar{y}_{\text{RC}}$) as given in (3). Moreover, for $\alpha=0,$ the proposed class T reduces to the exponential version of (Bahl and Tuteja, 1991) [1] estimator ($\bar{y}_{\exp}$) as given in (4). Hence $\bar{y}_{\text{RC}}$ and $\bar{y}_{\exp}$ can be regarded as the particular members of the proposed class T.

The mean square error (MSE) of the proposed class T is obtained on considering the following assumptions
$\bar{y}_{\text{st}}=\bar{Y}\left(1+e_0\right)$ and $\bar{x}_{\text{st}}=\bar{X}\left(1+e_1\right)$.

Then, we have

$$E\left(e_0^2\right)=\frac{\sum_{h=1}^{L}W_h^2\lambda_hS_{yh}^2}{\bar{Y}^2},\ E\left(e_1^2\right)=\frac{\sum_{h=1}^{L}W_h^2\lambda_hS_{xh}^2}{\bar{X}^2},\ E\left(e_0e_1\right)=\frac{\sum_{h=1}^{L}W_h^2\lambda_hS_{yxh}}{\bar{Y}\,\bar{X}}. \tag{13}$$

Now, expressing T in terms of e_0 and e_1, we have

$$T=\bar{Y}\left(1+e_0\right)\left[\alpha\left(1+e_1\right)^{-1}+\left(1-\alpha\right)\exp\left\{\left(\frac{-e_1}{2}\right)\left(1+\frac{e_1}{2}\right)^{-1}\right\}\right] \tag{14}$$

Expanding the R.H.S. of (14), multiplying out and retaining the first order error terms, we have

$$T = \bar{Y}\left[\alpha\left(1+e_0-e_1\right)+\left(1-\alpha\right)\left(1+e_0-\frac{e_1}{2}\right)\right] \tag{15}$$

$$\text{i.e., } T-\bar{Y} = \bar{Y}\left[\alpha\left(\frac{-e_1}{2}\right)+e_0-\frac{e_1}{2}\right] \tag{16}$$

Squaring both sides of (16), taking the expectation, and using the results of (13), we obtain the MSE of proposed class T to the first order of approximation as:

$$MSE(T) = \sum_{h=1}^{L} W_h^2\lambda_h\left[S_{yh}^2+\frac{(\alpha+1)^2}{4}R^2S_{xh}^2-(\alpha+1)RS_{yxh}\right] \tag{17}$$

The optimum value of α, which minimizes the MSE of proposed class T in (17), is obtained on using the following condition:

$$\frac{\partial}{\partial\alpha}MSE\left(T\right)=0 \tag{18}$$

On simplification, the optimum value of α is obtained as:

$$\alpha_{opt} = \frac{2\rho_{yxh}S_{yh}}{RS_{xh}}-1 = \frac{2\beta_h}{R}-1 \tag{19}$$

Where $\beta_h = \rho_{yxh}S_{yh}/S_{xh}$ denotes the population regression coefficient of Y on X in the h^{th} stratum.
On substituting the optimum value of α in (17), the minimum attainable MSE of the proposed class T is obtained as:

$$MSE(T)_{\min} = \sum_{h=1}^{L} W_h^2\lambda_h S_{yh}^2\left(1-\rho_{yxh}^2\right)=\text{MSE}\left(\bar{y}_{\text{lrs}}\right) \tag{20}$$

4. Efficiency Comparisons

The necessary and sufficient conditions (NASCs) for the superiority of the proposed class T as compared to the preliminary estimators are obtained on using the MSE criterion, by utilizing the equations (6) to (11), and (17) as follows:

$$MSE(T) < Var(\bar{y}_{st}) \text{ if } (\alpha+1) < \frac{4\beta_h}{R} \tag{21}$$

$$MSE(T) < MSE(\bar{y}_{RS}) \text{ if } \frac{(\alpha+1)^2}{4}R^2-\beta_h(\alpha+1)R < R_h^2-2\beta_hR_h \tag{22}$$

$$MSE(T) = MSE(\bar{y}_{lrs}) \text{ if } \alpha = \frac{2\beta_h}{R}-1 \tag{23}$$

$$MSE(T) < MSE(\bar{y}_{RC}) \text{ if } \frac{(\alpha+1)^2}{4}R^2-\beta_h(\alpha+1)R < R^2-2\beta_hR \tag{24}$$

$$MSE(T) < MSE(\bar{y}_{\exp}) \text{ if } \frac{(\alpha+1)^2}{4} R^2 - \beta_h (\alpha+1) R < \frac{R^2}{4} - \beta_h R \quad (25)$$

$$MSE(T) < MSE(\bar{y}_{lrc}) \text{ if } \frac{(\alpha+1)^2}{4} R^2 - \beta_h (\alpha+1) R < \frac{-\rho_{yx}^2 S_{yh}^2}{S_{xh}^2} \quad (26)$$

5. Empirical Analysis

To assess the efficiency of the proposed class T as compared to the preliminary estimators of the population mean ($\bar{Y}$), we have considered three real population datasets. The descriptions of the populations and the values of various parameters are given below:

Population I- (Singh and Chaudhary, 1988) [11]
Y : Total number of trees,
X : area under orchards in hectares.

Population II- (Kadilar and Cingi, 2005) [12]
Y : Amount of apple production,
X : Number of apple trees.

Population III – (Koyuncu and Kadilar, 2010) [13]
Y : Number of teachers,
X : Number of students.

Table 1: Parameters of Population I

Total	Stratum	1	2	3
$N = 25$ $n = 10$ $\bar{Y} = 410.841$ $\bar{X} = 8.3796$	N_h	6	8	11
	n_h	3	3	4
	$\bar{Y}_h$	417.33	503.38	340.00
	$\bar{X}_h$	6.81	10.12	7.97
	S_{yh}^2	74775.47	259113.70	65885.60
	S_{xh}^2	15.97	132.66	38.44
	S_{yxh}	1007.05	5709.16	1404.71

Table 2: Parameters of Population II

Total	Stratum	1	2	3	4	5	6
$N = 854$	N_h	106	106	94	171	204	173
$n = 140$	n_h	9	17	38	67	7	2
$\bar{Y} = 2930$	$\bar{Y}_h$	1536	2212	9384	5588	967	404
$\bar{X} = 37600$	$\bar{X}_h$	24375	27421	72409	74365	26441	9844
$R = 0.0779$	S_{yh}	6425	11532	29907	28643	2390	946
$\rho_{yx} = 0.8267$	S_{xh}	49189	57461	160757	285603	45403	18794
	ρ_{yxh}	0.82	0.86	0.90	0.99	0.71	0.89

Table 3: Parameters of Population III

Total	Stratum	1	2	3	4	5	6
$N=923$	N_h	127	117	103	170	205	201
$n=180$	n_h	31	21	29	38	22	39
$\bar{Y}=436.433$	$\bar{Y}_h$	703.74	413	573.17	424.66	267.03	393.84
$\bar{X}=11440.50$	$\bar{X}_h$	20804.59	9211.79	14309.30	9478.85	5569.95	12997.59
$R=0.0381$	S_{yh}	883.897	645.106	1033.43	810.676	403.749	711.669
$\rho_{yx}=0.9552$	S_{xh}	30478.70	15181	27545.40	18218.30	8499.74	23096.70
	ρ_{yxh}	0.936	0.996	0.994	0.983	0.989	0.965

The percent relative efficiencies (PREs) of the proposed class as well as the preliminary estimators are computed with respect to the stratified sample mean ($\bar{y}_{st}$) and the findings are elaborated in Table 4. The PREs are obtained on using the following formula:

$$\mathrm{PRE}\left(\phi,\bar{y}_{st}\right)=\frac{Var\left(\bar{y}_{st}\right)}{MSE\left(\phi\right)}\times 100,$$

Where $\phi=\bar{y}_{st},\bar{y}_{RS},\bar{y}_{lrs},\bar{y}_{RC},\bar{y}_{\exp},\bar{y}_{lrc},T.$

Table 4: PREs of various estimators of $\bar{Y}$

Estimator	Population		
	I	II	III
$\bar{y}_{st}$	100	100	100
$\bar{y}_{RS}$	815.55	422.65	1718.33
$\bar{y}_{lrs}$	982.04	623.49	2076.97
$\bar{y}_{RC}$	713.96	309.82	1025.32
$\bar{y}_{\exp}$	373.99	312.09	950.58
$\bar{y}_{lrc}$	872.06	315.89	1141.79
T	982.04	623.49	2076.97

Bold values signify the maximum PRE.

6. Results

The following results are obtained from Table 4:

1. In all three populations, the proposed class T has maximum PREs as compared to the stratified sample mean $(\bar{y}_{st})$, separate ratio estimator $(\bar{y}_{RS})$, combined ratio estimator $(\bar{y}_{RC})$, exponential ratio estimator $(\bar{y}_{\exp})$, and combined regression estimator $(\bar{y}_{lrc})$.
2. The proposed class T and the separate regression estimator $(\bar{y}_{lrs})$ are equally efficient for the estimation of population mean ($\bar{Y}$) of the study variable (Y), as was established theoretically in Section 3.
3. Among the specific members $\bar{y}_{RC}$ and $\bar{y}_{\exp}$ of the proposed class T, the estimator $\bar{y}_{RC}$ performs better than $\bar{y}_{\exp}$ in population I and III, except in population II.

7. Conclusion

In this paper, a composite class of estimators has been formulated for estimating the population mean of a study variable in stratified random sampling. It has been established that the estimators $\bar{y}_{RC}$ and $\bar{y}_{\exp}$ are specific members of the proposed class T. Moreover, the results of Table 4 reveal that the proposed class T is superior as compared to the stratified sample mean $(\bar{y}_{st})$, and the other preliminary estimators. Hence we conclude that the proposed class T has greater applicability for the estimation of mean in stratified random sampling. The proposed class T can also be regarded as an alternative to the separate regression estimator $(\bar{y}_{lrs})$.

8. References

1. Bahl S, Tuteja RK. Ratio and product type exponential estimators. Journal of Information and Optimization Sciences. 1991;12(1):159-164.
2. Upadhyaya LN, Singh HP. Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal. 1999;41:627-636.
3. Kadilar C, Cingi H. Ratio estimators in stratified random sampling. Biometrical journal. 2003;45(2):218–225.
4. Shabbir J, Gupta S. A new estimator of population mean in stratified sampling. Communications in Statistics-Theory and Methods. 2006;35(7):1201-1209.
5. Singh R, Kumar M, Chaudhary MK, Kadilar C. Improved exponential estimator in stratified random sampling. Pakistan Journal of Statistics and Operation Research. 2009;V(2):67-82.
6. Vishwakarma GK, Kumar M. Efficient classes of ratio-cum-product estimators of population mean in stratified random sampling. Probability and Mathematical Statistics. 2016;36(2):267-278.
7. Kumar M, Vishwakarma GK. Generalized classes of regression-cum-ratio estimators of population mean in stratified random sampling. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences. 2020;90(5):933-939.
8. Kumar M, Tiwari A. A generalized class of estimators of population mean in two-phase sampling using two auxiliary variables. International Journal of Mathematics and Statistics. 2021;22(3):18-27.
9. Kumar M, Tiwari A. A composite class of ratio estimators for a finite population mean in two-phase sampling. International Journal of Statistics and Reliability Engineering. 2022;9(3):364-370.
10. Tiwari A, Kumar M, Dubey SK. A generalized approach for estimation of a finite population mean in two-phase sampling. Asian Journal of Probability and Statistics. 2023;21(3):45-58.
11. Singh D, Chaudhary FS. Theory and Analysis of Sample Survey Designs, New Age International, New Delhi; c1986.
12. Kadilar C, Cingi H. A new ratio estimator in stratified random sampling. Communications in Statistics- Theory and Methods. 2005;34:597-602.
13. Koyuncu N, Kadilar C. On improvement in estimating population mean in stratified random sampling. Journal of Applied Statistics. 2010;37(6):999-1013.