International Journal of Statistics and Applied Mathematics 2023; 8(2): 90-93

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(2): 90-93

https://www.mathsjournal.com
Received: 10-01-2023
Accepted: 11-02-2023

Pooja Rawat
Department of Mathematics and Statistics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, Haryana, India

Manoj Kumar
Department of Mathematics and Statistics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, Haryana, India

Ajay Sharma
Department of Mathematics and Statistics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, Haryana, India

Corresponding Author:
Pooja Rawat
Department of Mathematics and Statistics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, Haryana, India

Estimation of population mean using auxiliary information: A simulation approach

Pooja Rawat, Manoj Kumar and Ajay Sharma

DOI: https://doi.org/10.22271/maths.2023.v8.i2b.957

Abstract
In this paper, different existing ratio estimators using one auxiliary variable are reviewed and their efficiencies are compared with known correlation coefficient. A bivariate population was generated using R software. Simple random sampling without replacement method was used for the selection of sample from the generated population. The different ratio estimators were compared with respect to bias, mean square error, skewness and kurtosis using simulation technique. Efficiencies of the estimators were compared and it is found that in comparison with traditional ratio estimator, all the estimators viz. E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8, E_9 and E_{10} were more efficient whereas E_{11} and E_{12} were less efficient for all the studied sample sizes and correlation coefficients. It was observed that as sample size increase then bias and mean square error both decreases within each correlation coefficient between X and Y. All the estimators were almost unbiased for both the populations for large sample size.

Keywords: Auxiliary variable, bivarite population, simulation, bias, mean square error

1. Introduction
Sample surveys are extensively used as a cost-effective apparatus of data collection for making valid inference about population parameters. In sample surveys, it is possible to measure certain characters other than the study character which are highly correlated with the study variable. This additional information obtained is known as auxiliary information. Several sample surveys were performed in India and abroad using auxiliary information which is highly correlated with the variable of interest. Tripathi (1978) [18] used the auxiliary information on one or more variables in sample surveys in three basic ways. At the pre-selection stage, at the selection stage (or design stage), at the estimation stage. In sample surveys, several authors like Singh and Solanki (2012) [13], Misra (2018) [7], Muhammad *et al.* (2019) [10], Kumar and Kumar (2020) [6], Ahuja *et al.* (2021) [2] and others have widely utilized auxiliary information in different forms in sample surveys to increase the performance of the estimators of the study variable. The ratio estimator usually performs well when there is a positive correlation between the study and auxiliary variables. Ratio method of estimation is further improved by Sisodia and Dwivedi (1981) [14], Bahl and Tuteja (1991) [3] Upadhyaya and Singh (1991) and Kadilar and Cingi (2004) [5]. Some notable works on various kind of ratio method of estimation are Subramani and Kumarapandiyan (2012), Abid *et al.* (2016) [1], Kanwai *et al.* (2016), Singh *et al.* (2019) [11], Singh and Yadav (2020) [12] and Tiwari *et al.* (2021) [17].
In statistics, it is very difficult task to obtain the real-life data for comparison of estimators under realistic conditions. Therefore, the concept of simulation will be used to generate data for the comparison of estimators under realistic conditions. Reddy *et al.* (2010) and Srinivas *et al.* (2013) had used concept of simulation for the comparison of ratio and ratio-cum-product estimators.
In this research paper, a simulation study had been done by generating the population for different values of correlation coefficient using bivariate normal distribution and studied properties of the estimators based on the generated population.

2. Material and Methods

Table 1: List of different ratio estimators of population mean with their bias and mean squared error

Estimator	Name	Bias	Mean squared error
$E_0 = \frac{\bar{y}}{\bar{x}}\bar{X}$	Classical Ratio estimators	$\frac{(1-f)}{n}\left(\frac{RS_x^2}{\bar{X}} - \frac{W\bar{Y}S_x^2}{\bar{X}^2}\right)$	$\frac{(1-f)}{n}\left(S_y^2 - 2WR^2S_x^2 + R^2S_x^2\right)$
$E_1 = \bar{y}\left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right)$	Sisodia and Dwivedi (1981) [14]	$\frac{(1-f)}{n}\left(\frac{R\phi_1^2S_x^2}{\bar{X}} - \frac{W\phi_1\bar{Y}S_x^2}{\bar{X}^2}\right)$	$\frac{(1-f)}{n}\left(S_y^2 - \phi_1WR^2S_x^2 + R^2\phi_1^2S_x^2\right)$
$E_2 = \bar{y}\exp\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]$	Bahl and Tuteja (1991) [3]	$\frac{(1-f)}{8n}\left(\frac{3RS_x^2}{\bar{X}} - \frac{4W\bar{Y}S_x^2}{\bar{X}^2}\right)$	$\frac{(1-f)}{n}\left(\frac{4S_y^2 + R^2S_x^2 - 4R^2S_x^2W}{4}\right)$
$E_3 = \bar{y}\left[\frac{C_x\bar{X}+\beta_2(x)}{C_x\bar{x}+\beta_2(x)}\right]$ $E_4 = \bar{y}[\frac{\beta_2(x)\bar{X}+C_x}{\beta_2(x)\bar{x}+C_x}]$	Upadhyaya and Singh (1999) [19]	$(\bar{y}_e) = \frac{(1-f)}{n}\left(\frac{R\phi_a^2S_x^2}{\bar{X}} - \frac{W\phi_a\bar{Y}S_x^2}{\bar{X}^2}\right)$ e=3,4 a=3,26	$(\bar{y}_e) = \frac{(1-f)}{n}\left(S_y^2 - \frac{2\phi_aWS_x\bar{Y}S_yS_xR}{S_y\bar{X}} + R^2\phi_a^2S_x^2\right)$
$E_5 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{\bar{x}}\bar{X}$ $E_6 = \frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+\beta_2(x))}[\bar{X}+\beta_2(x)]$	Kadilar and Cingi (2004) [5]	$(\bar{y}_f) = \frac{(1-f)}{n}\phi_{iK}^2\frac{S_x^2}{\bar{Y}}$ f= 5, 6 iK= 14, 15	$(\bar{y}_f) = \frac{(1-f)}{n}[\phi_{iK}^2S_x^2 + S_y^2(1-\rho^2)]$
$E_7 = \bar{y}\frac{\bar{X}+\rho}{\bar{x}+\rho}$	Singh and Tailor (2003)	$(\bar{y}_7) = \frac{(1-f)}{n}\left(\frac{R\phi_4^2S_x^2}{\bar{X}} - \frac{W\phi_4\bar{Y}S_x^2}{\bar{X}^2}\right)$	$(\bar{y}_{10}) = \frac{(1-f)}{n}\left(S_y^2 - 2\phi_4WR^2S_x^2 + R^2\phi_4^2S_x^2\right)$
$E_8 = \bar{y}[\frac{\bar{X}+\beta_1(x)}{\bar{x}+\beta_1(x)}]$ $E_9 = \bar{y}[\frac{C_x\bar{X}+\beta_1(x)}{C_x\bar{x}+\beta_1(x)}]$	Yan and Tian (2010)	$(\bar{y}_d) = \frac{(1-f)}{n}\left(\frac{R\phi_b^2S_x^2}{\bar{X}} - \frac{W\phi_b\bar{Y}S_x^2}{\bar{X}^2}\right)$ d=8,9 b=5, 8	$(\bar{y}_d) =$ λS 2 y (1 − ρ 2) $\left(S_y^2 - 2\phi_bWR^2S_x^2 + R^2\phi_b^2S_x^2\right)$
$E_{10} = \bar{y}\frac{\bar{X}+Q_a}{\bar{x}+Q_a}$	Subramani and Kumarapandiyan (2012)	$\bar{y}_{iJ}) = \frac{(1-f)}{n}\left(\frac{R\theta_i^2S_x^2}{\bar{X}} - \frac{W\theta_i\bar{Y}S_x^2}{\bar{X}^2}\right)$ i = 12, ij = 24	MSE $(\bar{y}_{ij}) = \frac{(1-f)}{n}\left(S_y^2 - 2\theta_iWR^2S_x^2 + R^2\theta_i^2S_x^2\right)$
$E_{11} = \bar{y}_r - \frac{\bar{y}}{n}\left[\frac{s_x^2}{\bar{x}\bar{X}} - \frac{s_{xy}}{\bar{y}\bar{x}}\right]$	Banti Kumar (2017)	0	$MSE(\bar{y}_r) + \frac{1}{n^2}(5C_{20}C_{11} - C_{20}^2 - 4C_{20}C_{02} + 2C_{11}^2)$

The following algorithm was used to generate correlated bivariate distributions is given by

- Generate two independent random variables U and V having the identical distribution.
- Set W = ρU+ $\sqrt{1-\rho^2}$ V
- Return the correlated pair (U, W)

The concept of Reddy *et al.* (2010) was used in the present investigation to generate population of size 1000 with various correlation coefficients i.e., 0.5 and 0.8. The details of the population are as follows:

The population have variance ratio $\frac{\sigma^2{}_x}{\sigma^2{}_y}$ = 1. This population will have the marginal distributions U ~ N (10,4) and W ~ N (10ρ + 10$\sqrt{1-\rho^2}$, 4), where ρ is the population correlation between U and W.

Different ratio estimators of the population mean were compared by generating the population of size 1000 using normal distribution. A sample of 500 of sizes n = 10, 25 and 50 were drawn and computed value of each of the estimators along with bias and mean square error.

3. Results and Discussion

Comparison of the studied estimators has been made in this section under the population with two correlated variables (U, W) having same variance. A bivariate population of size N = 1000 has been generated having ρ = 0.5 and 0.8. From the generated population, 500 simple random sample without replacement of size n = 10, 25 and 50 have been drawn to study the properties of various estimators.

Subcases

Case A: Comparison of studied ratio estimators of population mean having correlation coefficient ρ = 0.5 and sample size n = 10, 25 and 50.

Case B: Comparison of studied ratio estimators of population mean having correlation coefficient ρ = 0.8 and sample size n = 10, 25 and 50.

Case A: Comparison of studied ratio estimators of population mean having correlation coefficient ρ = 0.5 and at sample size n=10, 25 and 50

Table 3.1.2 provides the estimates of population mean obtained by different estimators along with its bias, MSE, RE, skewness and kurtosis. The correlation coefficient between auxiliary and study variable was 0.5 and sample size taken as n = 10, 25 and 50. It was concluded from the table that Upadhyaya and Singh (E_4, 1999) [19] was the best estimator in case of sample size n=10 whereas Bahl and Tuteja (E_2, 1991) [3] and Subramani and Kumarapandiyan (E_{10}, 2012) were most efficient for sample size 25 and 50. Kadilar and Cingi (E_5, 2004) [5] estimator was least efficient estimator for all sample sizes. It was also observed that estimators E_4 had negative value of bias for sample size 10 and 25. As the sample size increases from n=10 to n=50, then bias and mean square error both decreases. An attempt is also made to check the distribution pattern of studied estimators based on skewness and kurtosis and found that all the estimators were asymptotically normal.

Table 2: Comparison of different ratio estimators of population mean having correlation coefficient ρ=0.5

Estimator	Estimate	Bias	MSE	RE (E_i, E_0)	Skewness	Kurtosis
n = 10						
E_0	13.6679	0.0114	0.2204	-	0.4800	3.5676
E_1	13.6698	0.0110	0.2168	1.0166	0.4446	3.4855
E_2	13.6936	0.0033	0.1294	1.7032	0.1743	2.8981
E_3	13.6686	0.0113	0.2189	1.0069	0.4671	3.5370
E_4	13.6899	-0.0005	0.1284	1.7165	0.1880	2.9520
E_5	13.6672	0.0191	0.3886	0.5672	0.2061	3.0392
E_6	13.6787	0.0123	0.2960	0.7446	0.2963	3.1850
E_7	13.6703	0.0100	0.2062	1.0689	0.4349	3.4638
E_8	13.6682	0.0113	0.2196	1.0036	0.4734	3.5520
E_9	13.6688	0.0107	0.2133	1.0333	0.4633	3.5281
E_{10}	13.6937	0.0009	0.1294	1.7032	0.1639	2.8837
E_{11}	13.5233	0.0000	0.1938	1.1373	-0.1687	2.6079
n = 25						
E_0	13.7518	0.0020	0.0366	-	0.4374	3.4530
E_1	13.7550	0.0020	0.0362	1.0110	0.4254	3.4477
E_2	13.7955	0.0006	0.0201	1.8209	0.2765	3.3122
E_3	13.7530	0.0020	0.0365	1.0027	0.4331	3.4510
E_4	13.7891	-0.0001	0.0206	1.7767	0.3030	3.3451
E_5	13.7503	0.0032	0.0642	0.5701	0.0094	3.1935
E_6	13.7697	0.0020	0.0474	0.7722	0.3712	3.4218
E_7	13.7560	0.0018	0.0341	1.0733	0.4221	3.4463
E_8	13.7524	0.0020	0.0362	1.0110	0.4352	3.4519
E_9	13.7533	0.0015	0.0317	1.1546	0.4318	3.4504
E_{10}	13.7956	0.0002	0.0201	1.8209	0.2681	3.3045
E_{11}	13.6901	0.0000	0.0322	1.1366	-0.1687	2.6079
n = 50						
E_0	13.8331	0.0005	0.0089	-	0.1725	2.6287
E_1	13.8325	0.0005	0.0088	1.0114	0.1566	2.6219
E_2	13.8243	0.0001	0.0048	1.8542	0.0259	2.7446
E_3	13.8329	0.0005	0.0089	1.0000	0.1667	2.6258
E_4	13.8256	0.0000	0.0052	1.7115	0.0376	2.7228
E_5	13.8336	0.0008	0.0156	0.5705	0.0629	2.9299
E_6	13.8297	0.0005	0.0114	0.7807	0.0863	2.6421
E_7	13.8323	0.0004	0.0083	1.0723	0.1521	2.6207
E_8	13.8330	0.0005	0.0088	1.0114	0.1695	2.6272
E_9	13.8328	0.0004	0.0075	1.1867	0.1650	2.6251
E_{10}	13.8243	0.0000	0.0048	1.8542	0.0294	2.7436
E_{11}	13.8030	0.0000	0.0078	1.1410	-0.1687	2.6079

Case: B Comparison of studied ratio estimators of population mean having correlation coefficient ρ = 0.8 and at sample sizes n=10, 25 and 50

Table 3.1.4 provides the estimates of population mean obtained by different estimators along with its bias, MSE, RE, skewness and kurtosis. The correlation coefficient between auxiliary and study variable was 0.8 and sample size taken as n = 10, 25 and 50. Among the studied estimators Bahl and Tuteja (E_2, 1991) [3] and Subramani and Kumarapandiyan (E_{10}, 2012) were the most efficient and Kadilar and Cingi (E_6, 2004) [5] was the least efficient estimator. It was also observed that estimators E_4 had negative value of bias and for all sample sizes. An attempt is also made to check the distribution pattern of studied estimators based on skewness and kurtosis and found that all the estimators were asymptotically normal.

Table 3: Comparison of different ratio estimators of population mean having correlation coefficient ρ=0.8

Estimator	Estimate	Bias	MSE	RE (E_i, E_0)	Skewness	Kurtosis
n = 10						
E_0	14.0340	0.0077	0.1001	-	0.4607	3.6901
E_1	14.0348	0.0073	0.0976	1.0256	0.4064	3.5493
E_2	14.0436	0.0014	0.0616	1.6250	0.0887	2.8315
E_3	14.0343	0.0076	0.0990	1.0111	0.4406	3.6356
E_4	14.0421	-0.0018	0.0806	1.2419	0.0918	2.8961
E_5	14.0337	0.0197	0.3351	0.2987	0.1979	3.1961
E_6	14.0379	0.0127	0.2367	0.4229	0.1992	3.1769
E_7	14.0355	0.0057	0.0856	1.1694	0.3566	3.4388
E_8	14.0342	0.0076	0.0998	1.0030	0.4498	3.6600
E_9	14.0344	0.0074	0.0982	1.0193	0.4331	3.6158
E_{10}	14.0436	-0.0011	0.0616	1.6250	0.0760	2.8144
E_{11}	13.8894	0.0000	0.1102	0.9083	-0.1687	2.6079
n = 25						
E_0	14.0731	0.0014	0.0172	-	0.3648	3.2573
E_1	14.0764	0.0013	0.0169	1.0178	0.3481	3.2569
E_2	14.1179	0.0003	0.0096	1.7917	0.1661	3.2006

E_3	14.0743	0.0014	0.0171	1.0058	0.3585	3.2562
E_4	14.1109	-0.0003	0.0155	1.1097	0.2158	3.2547
E_5	14.0706	0.0033	0.0566	0.3039	-0.0667	3.2031
E_6	14.0903	0.0021	0.0387	0.4444	0.2957	3.3176
E_7	14.0797	0.0010	0.0146	1.1781	0.3339	3.2638
E_8	14.0738	0.0013	0.0170	1.0118	0.3613	3.2566
E_9	14.0748	0.0010	0.0145	1.1862	0.3562	3.2561
E_{10}	14.1180	-0.0001	0.0096	1.7917	0.1500	3.2191
E_{11}	14.0112	0.0000	0.0191	0.9005	-0.1687	2.6079
n = 50						
E_0	14.1352	0.0003	0.0042	-	0.2114	2.7122
E_1	14.1348	0.0003	0.0041	1.0244	0.1849	2.6682
E_2	14.1299	0.0001	0.0023	1.8261	-0.0058	2.7653
E_3	14.1351	0.0003	0.0042	1.0000	0.2018	2.6955
E_4	14.1307	-0.0001	0.0042	1.0000	-0.0014	2.7370
E_5	14.1358	0.0008	0.0137	0.3066	0.0578	3.0747
E_6	14.1334	0.0005	0.0093	0.4516	0.0639	2.5993
E_7	14.1344	0.0003	0.0036	1.1667	0.1589	2.6326
E_8	14.1352	0.0003	0.0041	1.0244	0.2062	2.7030
E_9	14.1350	0.0002	0.0034	1.2353	0.1982	2.6893
E_{10}	14.1299	0.0000	0.0023	1.8261	-0.0049	2.7582
E_{11}	14.1052	0.0000	0.0047	0.8936	-0.1687	2.6079

Conclusion

From the study, it is observed that the efficiency of the different estimator changes with the values of the correlation coefficient as well as sample sizes. The best estimator in terms of mean square error value for different dimension is given below in the table.

- It is observed that as sample size increase then bias and mean square error both decreases within each correlation coefficient between X and Y.
- As the correlation coefficient increases the value of Mean Square Error decreases.
- As the sample size increases the value of bias becomes almost constant.
- All the estimators are almost unbiased for both the populations for large sample size.
- From the skewness and kurtosis values, it is found that all the ratio estimators are asymptotically normal.

References

1. Abid M, Abbas N, Nazir HZ, Lin Z. Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. Revista Colombiana de Estadistica. 2016;39(1):63-79.
2. Ahuja TK, Misra P, Behwal OK. A generalized two-phase sampling estimator of ratio of population means using auxiliary information. Journal of Reliability and Statistical Studies. 2021;14(1):1-16.
3. Bahl S, Tuteja RK. Ratio and product type exponential estimators. Journal of Information and Optimaztion Sciences. 1991;12(1):159-164.
4. Bulut H, Zaman T. An improved class of robust ratio estimators by using the minimum covariance determinant estimation. Communications in Statistics-Simulation and Computation. 2022;51(5):2457-2463.
5. Kadilar C, Cingi H. Ratio estimators in simple random sampling. Applied Mathematics and Computation. 2004;151(3):893-902.
6. Kumar K, Kumar S. Two phase sampling exponential type estimators for population mean using auxiliary attribute in the presence of non-response. International Journal of Mathematics and Statistics. 2020;21(1):75-85.
7. Misra P. Regression type double sampling estimator of population mean using auxiliary information. Journal of Reliability and Statistical Studies. 2018;11(1):21-28.
8. Mittal A, Kumar M. Generalised exponential ratio-type estimator for finite population variance under random non-response. International journal of Advanced Research. 2021;9(01):589-596.
9. Mradula SK, Yadav R Varshney, Dube M. Efficient estimation of population mean under stratified random sampling with linear cost function. Communications in Statistics-Simulation and Computation. Advance online publication; c2019.
10. Muhammad I, Maria J, Zhengyan L. Enhanced estimation of population mean in the presence of auxiliary information, Journal of King Saud University-Science. 2019;31(4):1373-1378.
11. Singh A, Vishwakarma GK, Gangele RK. Improved predictive estimators for finite population mean using Searls technique. Journal of Statistics and Management Systems. 2019;22(8):1555-1571.
12. Singh HP, Yadav A. A new exponential approach for reducing the mean squared errors of the estimators of population mean using conventional and non-conventional location parameters. Journal of Modern Applied Statistical Methods. 2020;18(1):1-46.
13. Singh HP, Solanki RS. An alternative procedure for estimating the population mean in simple random sampling. Pakistan Journal of Statistics and Operation Research. 2012;8(2):213-232.
14. Sisodia BVS, Dwivedi VK. A modified ratio estimator using coefficient of variation of auxiliary variable. Journal of the Indian Society of Agriculture Statistics. 1981;33(1):13-18.
15. Subramani J, Kumarapandiyan G. Estimation of Population Mean using Known Correlation Coefficient and Median. Journal of Statistical Theory and Applications. 2014;13(4):333-343.
16. Kumar TM, Jaslam PKM. Efficiency evaluation of ratio estimators in simple random sampling. Journal of Reliability and Statistical Studies. 2019;12(2):103-113.
17. Tiwari KK, Bhougal S, Kumar S. A Difference-Cum-Exponential Type Efficient Estimator of Population Mean. In Mathematical Modeling and Computation of Real-Time Problems. CRC Press; c2021. p. 181-194.
18. Tripathi TP. A note on optimum weights in multivariate ratio, product and regression estimators. Journal of Indian Society of Agricultural Statistics. 1978;30(1):101-109.
19. Upadhyaya LN, Singh HP. Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal. 1999;41(5):627-636.