Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229 5666 Vol. 11, Issue 1 (2018): 83-92

ESTIMATION OF FINITE POPULATION MEAN IN STRATIFIED RANDOM SAMPLING USING NON-CONVENTIONAL MEASURES OF DISPERSION

[1]Mir Subzar*, S. Maqbool, T. A. Raja and M. A. Bhat
[1]Division of Agricultural Statistics, SKUAST-Kashmir, India
E Mail: [1]subzarstat@gmail.com

Received October 29, 2017
Modified May 15, 2018
Accepted June 10, 2018

Abstract

The present study was taken into consideration to suggest a proficient class of estimators for predetermined population mean of variable of interest in stratified random sampling by utilizing the auxiliary information of robust measures such as Gini's Mean Difference, Downton's Method and Probability Weighted Moments. Asymptotic properties such as bias and mean square error of the proposed class of estimators have been derived using Taylor series method upto first degree of approximation. In the support of the theoretical proposed work we have given numerical illustration and from this we conclude that our proposed class of estimators performs better than existing estimators.

Key Words: Non-Conventional Measures of Dispersion; Ratio Estimators; Stratified Random Sampling; Mean Square Error; Efficiency.

AMS Subject Classification: 62D05

1. Introduction

As we know that in sample survey, it is always advantageous to use the information accessible on the ancillary variable which is highly interrelated with the study variable. The use of auxiliary information increases the precision of the estimators used for estimating the unknown population parameters. Several authors have used auxiliary information on auxiliary variable in the estimation of population parameters like Srivastava and Jhajj (1981), Singh and Vishwakarma (2007), Sahai and Ray (1980),Bahl and Tuteja (1991), Srivastava and Jhajj (1983), Srivastava (1971), Swain (1970) and Perri (2007).

Here we have tried to incorporate the use of auxiliary information in stratified random sampling. Stratified random sampling is alternative method by which select the sample form the population. By this sampling design the sample from the heterogeneous population is selected in such a way that it represents the whole population. However, in this method the population data are grouped in strata and sample sizes in each stratum are randomly selected. Several authors like Kadilar and Cingi (2003, 2005), Haq and Shabbir (2013), Sangngam and Hiriote (2014), Shabbir and Gupta (2006), Kadilar and Cingi (2003), Maqbool et al. (2017a) and Maqbool et al. (2017b) have proposed the two ratio estimators in two separate papers by using the ancillary information of Mid-range and Hodges Lehmann respectively. Also Subzar et al. (2017) have also proposed some modified ratio estimators using ancillary information of decile mean, quartile deviation, median with correlation coefficient, coefficient of variation and coefficient of skewness in SRSWOR for population mean. So in the present study we have also utilized the auxiliary information of non-conventional location measures of dispersion in stratified sampling. We have used non-conventional measures of dispersion; as these measures are more robust against the outliers, present in the data.

Consider a predetermined population, $u = (u_1, u_2, ..., u_N)$. consists of different and detectible units which are assorted from each other. After that L strata of sizes, N_h, $(h = 1,2,...,L)$, are made of this population and the strata are homogeneous within and heterogeneous between. The study variable, y, and auxiliary variables, x, take the values, y_{hi} and x_{hi}, respectively, for the i^{th} unit of the h^{th} stratum. Form each strata we draw the samples of required size n_h ($h = 1,2,...,L$), using proportional allocation as $n = \sum_{h=1}^{L} n_h$. Before discussing about the existing estimators and proposed estimators, we will mention the notations which we have used in the present study and are given below

Notations

N	Population size
n_h	Sample size in stratum h
N_h	No. of units in stratum h
$W_h = \frac{N_h}{N}$	Stratum weight
C_{xh}	Coefficient of variation of stratum h,
β_{x2h}	Coefficient of kurtosis of stratum h,
l	No. of stratum
β_{x1h}	Coefficient of skewness of stratum h
$\bar{y}_h$	Sample mean of the study variable in stratum h,
$\gamma_h = (1 - f_h)/n_h$, $f_h = \frac{n_h}{N_h}$	sampling fraction in stratum h
S_{yh}^2	Variance of the study variable in stratum h,
$\bar{x}_{st} = \sum_{h=1}^{l} W_h \bar{x}_h$	Is an unbiased estimator of $\bar{X}$
$\bar{x}_h$	Sample mean of auxiliary variable in stratum h
S_{xh}^2	Variance of the auxiliary variable in stratum h

S_{xyh} Covariance between auxiliary and study variables in stratum

G_h Gini's mean difference of the stratum h

D_h Downton's method of the stratum h

$S_{pw(h)}$ Probability weighted moments of the stratum h

2. Existing Estimators in Stratified Sampling

In this section, we mention the existing ratio estimators in stratified random sampling given by different authors and are given as follows

Separate ratio estimator in stratified sampling

$$\bar{y}_{RS} = \sum_{h=1}^{l} W_h \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h . \text{(Kadilar \& Cingi, 2003)} \quad (2.1)$$

An approximated MSE and Bias of the separate ratio estimator given in (2.1) is given as

$$MSE(\bar{y}_{RS}) = \sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{RS}^2 S_{xh}^2 - 2R_{RS} S_{xyh})$$

Where $R_{RS} = \bar{Y}_h / \bar{X}_h$

$$Bias(\bar{y}_{RS}) = \sum_{h=1}^{l} W_h^2 \gamma_h \left(\frac{R_{RS}}{\bar{X}_h} S_{xh}^2 - \frac{1}{\bar{X}_h} S_{xyh} \right)$$

Combined ratio estimator in stratified sampling

$$\bar{y}_{RC\_KC} = \bar{y}_{st} \frac{\sum_{h=1}^{l} W_h (\bar{X}_h + C_{xh})}{\sum_{h=1}^{l} W_h (\bar{x}_h + C_{xh})} \quad \text{(Kadilar and Cingi, 2005)} \quad (2.2)$$

An approximated MSE and Bias of the separate ratio estimator given in (2.2) is given as

$$Bias(\bar{y}_{RC\_KC}) = \sum_{h=1}^{l} W_h^2 \gamma_h \left(\frac{R_{KC}}{\bar{X}_{KC}} S_{xh}^2 - \frac{1}{\bar{X}_{KC}} S_{xyh} \right)$$

$$MSE(\bar{y}_{RC\_KC}) = \sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{KC}^2 S_{xh}^2 - 2R_{KC} S_{xyh})$$

Where $R_{KC} = \frac{\bar{Y}_{st}}{\bar{X}_{KC}} = \frac{\bar{Y}_{st}}{\sum_{h=1}^{l} W_h (\bar{X}_h + C_{xh})}$

Combined ratio estimator in stratified sampling, when coefficient of variation is known

$$\bar{y}_{RC\_SD} = \frac{\bar{y}_{st}}{(\bar{x}_{st} + C_x)} (\bar{X} + C_x) \quad \text{(Sangngam and Hiriote, 2014)} \quad (2.3)$$

An approximated MSE and Bias of the separate ratio estimator given in (2.1) is given as

$$Bias(\bar{y}_{RC\_SD}) = \sum_{h=1}^{l} W_h^2 \gamma_h \left(\frac{R_{SD}}{\bar{X}_x + C_x} S_{xh}^2 - \frac{1}{\bar{X}_x + C_x} S_{xyh} \right)$$

$$MSE(\bar{y}_{RC\_SD}) = \sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{SD}^2 S_{xh}^2 - 2R_{SD} S_{xyh})$$

Where $R_{SD} = \dfrac{\bar{Y}}{\bar{X} + C_x}$

As the above estimators use the auxiliary information of population mean and coefficient of variation, which are affected by the presence of the outliers in the data. So keeping this problem under consideration, in this paper we have incorporated the auxiliary information of Non-Conventional Measures of dispersion which are robust against outliers. Thus our suggested estimators would be always superior to the above estimators for any data set.

3. Suggested Estimators

Keeping the above strategy in view we have advocated the novel modified ratio type estimators in stratified random sampling using the supplementary information of robust measures such as Gini's mean difference, Downton's method and probability weighted moments which are insensitive against outliers present in the data and the estimators with the expressions of bias and mean square error are given as under

$$\bar{y}_{pr1} = \frac{\bar{y}_{st}}{(\bar{x}_{st} + G_h)} (\bar{X} + G_h) \tag{3.1}$$

Where the Bias and mean square error of the equation (3.1) are given as under

$$Bias(\bar{y}_{pr1}) = \sum_{h=1}^{l} W_h^2 \gamma_h \left(\frac{R_{pr1}}{\bar{X} + G_h} S_{xh}^2 - \frac{S_{xyh}}{\bar{X} + G_h} \right)$$

$$MSE(\bar{y}_{pr1}) = \sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{pr1}^2 S_{xh}^2 - 2R_{pr1} S_{xyh})$$

Where $R_{pr1} = \dfrac{\bar{Y}_h}{(\bar{X}_h + G_h)}$

$$\bar{y}_{pr2} = \frac{\bar{y}_{st}}{(\bar{x}_{st} + D_h)} (\bar{X} + D_h) \tag{3.2}$$

Where the Bias and mean square error of the equation (3.2) are given as under

$$Bias(\bar{y}_{pr2}) = \sum_{h=1}^{l} W_h^2 \gamma_h \left(\frac{R_{pr2}}{\bar{X} + D_h} S_{xh}^2 - \frac{S_{xyh}}{\bar{X} + D_h} \right)$$

$$MSE(\bar{y}_{pr2}) = \sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{pr2}^2 S_{xh}^2 - 2R_{pr2} S_{xyh})$$

Where $R_{pr2} = \dfrac{\bar{Y}_h}{(\bar{X}_h + D_h)}$.

$$\bar{y}_{pr3} = \frac{\bar{y}_{st}}{(\bar{x}_{st} + S_{pw(h)})}(\bar{X} + S_{pw(h)}). \quad (3.3)$$

Where the Bias and mean square error of the equation (3.3) are given as under

$$Bias(\bar{y}_{pr3}) = \sum_{h=1}^{l} W_h^2 \gamma_h \left(\frac{R_{pr3}}{\bar{X} + S_{pw(h)}} S_{xh}^2 - \frac{S_{xyh}}{\bar{X} + S_{pw(h)}} \right)$$

$$MSE(\bar{y}_{pr3}) = \sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{pr3}^2 S_{xh}^2 - 2R_{pr3} S_{xyh})$$

Where $R_{pr3} = \dfrac{\bar{Y}_h}{(\bar{X}_h + S_{pw(h)})}$

4. Efficiency comparison

We compare here the suggested estimators with the existing estimators mentioned above. We will have the conditions as follows:

4.1 Comparison of suggested estimators with separate ratio estimator given by Kadilar & Cingi, 2003

$$MSE(\bar{y}_{pri}) < MSE(\bar{y}_{stRS})$$

$$\sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{pri}^2 S_{xh}^2 - 2R_{pri} S_{xyh}) <$$

$$\sum_{h=1}^{l} W_h^2 \gamma_h (S_{yh}^2 + R_{RS}^2 S_{xh}^2 - 2R_{RS} S_{xyh}) \quad (4.1.1)$$

$$\sum_{h=1}^{l} W_h^2 \gamma_h (R_{pri}^2 S_{xh}^2 - 2R_{pri} S_{xyh}) < \sum_{h=1}^{l} W_h^2 \gamma_h (R_{RS}^2 S_{xh}^2 - 2R_{RS} S_{xyh})$$

Let $A = \sum_{h=1}^{l} W_h^2 \gamma_h S_{xyh}$ and $B = \sum_{h=1}^{l} W_h^2 \gamma_h S_{xh}^2$

Then (4.1.1) becomes

$$-2R_{pri} A + R_{pri}^2 B < -2R_{RS} A + R_{RS}^2 B$$

$$-2R_{pri} A + 2R_{RS} A + R_{pri}^2 B - R_{RS}^2 B < 0$$

$$-2A(R_{pri} - 2R_{RS}) + B(R_{pri}^2 - R_{RS}^2) < 0$$

$$-2A(R_{pri} - 2R_{RS}) + B(R_{pri} - R_{RS})(R_{pri} + R_{RS}) < 0$$

Where there are two conditions as follows

(i) When $(R_{pri} - R_{RS})(R_{pri} + R_{RS}) < 0$

$$\frac{-2A}{R_{pri}+R_{RS}}+B<0$$

$$B<\frac{2A}{R_{pri}+R_{RS}}$$

(ii) When $(R_{pri}-R_{RS})(R_{pri}+R_{RS})>0$

$$\frac{-2A}{R_{pri}+R_{RS}}+B>0$$

$$B>\frac{2A}{R_{pri}+R_{RS}} \qquad \text{Where } i=1,2,3.$$

When the condition I and II are satisfied then our advocated estimators are more proficient than the estimators given by Kadilar and Cingi (2003).

4.2 Comparison of suggested estimators with combined ratio estimator in stratified sampling given by Kadilar & Cingi, 2005

$$MSE(\bar{y}_{pri})<MSE(\bar{y}_{stRS})$$

$$\sum_{h=1}^{l}W_h^2\gamma_h(S_{yh}^2+R_{pri}^2S_{xh}^2-2R_{pri}S_{xyh})<\sum_{h=1}^{l}W_h^2\gamma_h(S_{yh}^2+R_{KC}^2S_{xh}^2-2R_{KC}S_{xyh}) \tag{4.2.1}$$

$$\sum_{h=1}^{l}W_h^2\gamma_h(R_{pri}^2S_{xh}^2-2R_{pri}S_{xyh})<\sum_{h=1}^{l}W_h^2\gamma_h(R_{KC}^2S_{xh}^2-2R_{KC}S_{xyh})$$

Let $A=\sum_{h=1}^{l}W_h^2\gamma_hS_{xyh}$ and $B=\sum_{h=1}^{l}W_h^2\gamma_hS_{xh}^2$

Then (4.2.1) becomes

$$-2R_{pri}A+R_{pri}^2B<-2R_{KC}A+R_{KC}^2B$$

$$-2R_{pri}A+2R_{KC}A+R_{pri}^2B-R_{KC}^2B<0$$

$$-2A(R_{pri}-2R_{KC})+B(R_{pri}^2-R_{KC}^2)<0$$

$$-2A(R_{pri}-2R_{KC})+B(R_{pri}-R_{KC})(R_{pri}+R_{KC})<0$$

Where there are two conditions as follows

(i) When $(R_{pri}-R_{KC})(R_{pri}+R_{KC})<0$

$$\frac{-2A}{R_{pri}+R_{KC}}+B<0 \text{ or } B<\frac{2A}{R_{pri}+R_{KC}}$$

(ii) When $(R_{pri}-R_{KC})(R_{pri}+R_{KC})>0$

$$\frac{-2A}{R_{pri}+R_{KC}}+B>0$$
$$B>\frac{2A}{R_{pri}+R_{KC}}$$

When the condition I and II are satisfied then our advocated estimators are more proficient than the estimators given by Kadilar and Cingi (2005).

4.3 Comparison of suggested estimator with Combined ratio estimator in stratified sampling, when coefficient of variation is known, given by Sangngam and Hiriote (2014)

$$MSE(\bar{y}_{pri})<MSE(\bar{y}_{stSD})$$
$$\sum_{h=1}^{l}W_h^2\gamma_h(S_{yh}^2+R_{pri}^2S_{xh}^2-2R_{pri}S_{xyh})<$$
$$\sum_{h=1}^{l}W_h^2\gamma_h(S_{yh}^2+R_{SD}^2S_{xh}^2-2R_{SD}S_{xyh}) \qquad (4.3.1)$$
$$\sum_{h=1}^{l}W_h^2\gamma_h(R_{pri}^2S_{xh}^2-2R_{pri}S_{xyh})<\sum_{h=1}^{l}W_h^2\gamma_h(R_{SD}^2S_{xh}^2-2R_{SD}S_{xyh})$$

Let $A=\sum_{h=1}^{l}W_h^2\gamma_hS_{xyh}$ and $B=\sum_{h=1}^{l}W_h^2\gamma_hS_{xh}^2$

Then (4.3.1) becomes

$$-2R_{pri}A+R_{pri}^2B<-2R_{SD}A+R_{SD}^2B$$
$$-2R_{pri}A+2R_{SD}A+R_{pri}^2B-R_{SD}^2B<0$$
$$-2A(R_{pri}-2R_{SD})+B(R_{pri}^2-R_{SD}^2)<0$$
$$-2A(R_{pri}-2R_{SD})+B(R_{pri}-R_{SD})(R_{pri}+R_{SD})<0$$

Where there are two conditions as follows

(i) When $(R_{pri}-R_{SD})(R_{pri}+R_{SD})<0$

$$\frac{-2A}{R_{pri}+R_{SD}}+B<0$$
$$B<\frac{2A}{R_{pri}+R_{SD}}$$

(ii) When $(R_{pri}-R_{SD})(R_{pri}+R_{SD})>0$

$$\frac{-2A}{R_{pri}+R_{SD}}+B>0 \text{ or } B>\frac{2A}{R_{pri}+R_{SD}}$$

When the condition I and II are satisfied then our advocated estimators are more proficient than the estimators given by Sangngam and Hiriote (2014).

5. Empirical study

For the population, we use the data of cultivation and production of apple in district Baramulla of Kashmir in which the apple production (in tons) is denoted by Y (study variable) and number of apple trees are denoted by X (auxiliary variable, 1 unit = 100 trees) in 499 villages of the Baramulla region of Jammu and Kashmir in 2010-2011 (Source: RCM project, pilot survey for estimation of cultivation and production of apple in district Baramulla, RCM approved project). We apply the proposed and existing estimators to this data set and the data statistics of this population is given in Table 1. Statistical analysis of these estimators is given in Table 2 and percent relative efficiency is given in Table 3.

Stratum 1	**Stratum 2**	**Stratum 3**	**Stratum 4**
$N_1 = 156$	$N_2 = 116$	$N_3 = 117$	$N_4 = 110$
$n_1 = 15$	$n_2 = 15$	$n_3 = 15$	$n_4 = 15$
$\overline{X}_1 = 10.317$	$\overline{X}_2 = 12.117$	$\overline{X}_3 = 6.053$	$\overline{X}_4 = 7.187$
$\overline{Y}_1 = 22.15$	$\overline{Y}_2 = 37.97$	$\overline{Y}_3 = 21.79$	$\overline{Y}_4 = 25.94$
$\beta_{x11} = 0.12$	$\beta_{x12} = 0.10$	$\beta_{x13} = 0.29$	$\beta_{x14} = 0.31$
$\beta_{x21} = 1.15$	$\beta_{x22} = 1.01$	$\beta_{x23} = 1.24$	$\beta_{x24} = 1.32$
$C_{x1} = 28.69$	$C_{x2} = 19.72$	$C_{x3} = 39.54$	$C_{x4} = 48.78$
$C_{y1} = 86.02$	$C_{y2} = 57.99$	$C_{y3} = 39.54$	$C_{y4} = 48.11$
$S_{x1} = 2.960$	$S_{x2} = 2.389$	$S_{x3} = 2.393$	$S_{x4} = 3.506$
$S_{y1} = 19.06$	$S_{y2} = 22.02$	$S_{y3} = 8.62$	$S_{y4} = 12.48$
$\rho_1 = 0.840$	$\rho_2 = 0.860$	$\rho_3 = 0.890$	$\rho_4 = 0.901$
$G_1 = 9.678$	$G_2 = 10.338$	$G_3 = 7.345$	$G_4 = 7.231$
$D_1 = 8.678$	$D_2 = 9.637$	$D_3 = 6.889$	$D_4 = 6.345$
$S_{pw1} = 7.899$	$S_{pw2} = 8.678$	$S_{pw3} = 5.781$	$S_{pw4} = 7.678$

Table 1: Descriptive statistics of the population

Estimators	Bias	Mean square error
$\overline{y}_{RS}$	1.674860919	558.9088144
$\overline{y}_{RC\_KC}$	-0.147656756	438.1938213
$\overline{y}_{RC\_SD}$	-0.176567564	425.1938215
$\overline{y}_{pr1}$	-0.315678912	165.3324190
$\overline{y}_{pr2}$	-0.289675325	158.7867542
$\overline{y}_{pr3}$	-0.267897534	152.7875649

Table 2: Statistical analysis of the estimators for the population

	$\overline{y}_{RS}$	$\overline{y}_{RC\_KC}$	$\overline{y}_{RC\_SD}$
$\overline{y}_{pr1}$	338.0516	265.0381	257.1751
$\overline{y}_{pr2}$	351.9871	275.9637	267.7766
$\overline{y}_{pr3}$	365.8078	286.7994	278.2909

Table 3: Percent relative efficiency of the proposed estimators with the existing estimators

6. Conclusion

Thus from the efficiency comparison and numerical illustration of suggested estimators given in section 4 and section 5 respectively with the mentioned estimators in literature, we reveal that our estimators perform better than the mentioned estimators in literature as their mean square error and bias are lower than the mentioned estimators in literature. So keeping the above results in view, we then strongly recommend that our suggested estimators are not only robust but also proficient than the estimators in literature and preferred over estimators in literature for use in practical applications.

Acknowledgement

The authors are extremely obliged to the adjudicators and the Editor in chief for appreciated recommendations on original form of the Document.

References

1. Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimators, Journal of Information and Optimization Sciences, 12(1), p. 159-164.
2. Haq, A. and Shabbir, J. (2013). Improved family of ratio estimators in simple and stratified random sampling, Communications in Statistics- Theory and Methods, 42(5), p. 782-799.
3. Kadilar, C. and Cingi, H. (2003). Ratio estimators in stratified random sampling, Biometrical Journal, 45(2), p. 218-225.

4. Kadilar, C. and Cingi, H. (2005). A new Ratio estimators in stratified random sampling, Communication in Statistics – Theory and Methods, 34, p. 597-620.
5. Maqbool, S., Subzar, M. and Bhat, M. A. (2017a). Ratio estimator in stratified random sampling using mid-range as auxiliary information, International Journal for Research Trends and Innovation, 2(8), p. 15-18.
6. Maqbool, S., Subzar, M. and Bhat, M. A. (2017b). Ratio estimator in stratified random sampling using Hodges Lehmann as auxiliary information, International Journal of Novel Research and Development, 2(8), p. 1-3.
7. Perri, G. D. P. F. (2007). Estimation of finite population mean using multi-auxiliary information, Metron, 65(1), p. 99-112.
8. Sahai, A. and Ray, S. K. (1980). An efficient estimator using auxiliary information, Metrika, 27(4), p. 271-275.
9. Sangngam, P. and Hiriote, S. (2014). Modified ratio estimators in stratified random sampling, J. Sci. Technol. MSU,33(2), p. 112-116.
10. Shabbir, J. and Gupta, S. (2006). A new estimator of population mean in stratified sampling, Communications in Statistics- Theory and Methods, 35(7), p. 1201-1209.
11. Singh, H. P. and Vishwakarma, G. K. (2007). Modified exponential ratio and product estimators for finite population mean in double sampling, Australian Journal of Statistics, 36(3), p. 217-225.
12. Srivastava, S. K. (1971). A generalized estimator for the mean of a finite population using multi auxiliary information, Journal of the American Statistical Association, 66(334), p. 404-407.
13. Srivastava, S. K. and Jhajj, H. S. (1981). A class of estimators of the population mean in survey sampling using auxiliary information, Biometrika, 68(1), p. 341-343.
14. Srivastava, S. K. and Jhajj, H. S. (1983). A class estimators of the population means using multi auxiliary information, Calc. Statist. Assoc. Bull., 32 p. 47-56.
15. Subzar, M., Maqbool, S., Verma, M. R. and Raja, T. A. (2017). An improved class of ratio estimators for estimating population means using auxiliary information in survey sampling, Journal of Reliability and Statistical Studies, 10(2), p. 65-82.
16. Swain, A. K. (1970). A note on the use of multiple auxiliary variables in sample surveys, Trabajos de Estadistica y de Investigacion operative, 21(3), p. 135-141.