IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Use of Median in Calibration Estimation of the Finite Population Mean in Stratified Sampling

Neha Garg [1*], Menakshi Pachori [2]

[1,2] School of Sciences, Indira Gandhi National Open University, New Delhi-110068, India

[*]Corresponding Author: neha1garg@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i5.667671 | Available online at: www.ijcseonline.org

Accepted: 23/May/2019, Published: 31/May/2019

***Abstract*—** This paper proposes a new calibration estimator for estimating the finite population mean in stratified random sampling using a calibration constraint which consider known median of the auxiliary variable. The result has been extended in case of stratified double sampling when median of the auxiliary variable is not known. The efficiency of the proposed estimator has also been compared with the help of simulation study on a real dataset.

***Keywords*—** Auxiliary information, Calibration estimation, Median, Stratified Sampling, Double sampling.

I. INTRODUCTION

Calibration estimation is a method of adjusting weights in survey sampling in order to estimate population parameters with the help of auxiliary information. The calibration approach became more popular in recent past years. The calibration estimation is a procedure of minimizing a distance function subject to calibration constraints using prior information on various parameters related to auxiliary variables such as mean, standard deviation, etc. to obtain the more precise estimators pertaining to variable under study [1]. The contribution in developing the calibrated estimators for different population parameters using different calibration constraints under many sampling schemes are due to many researchers such as [2], [3], [4], [5], [6], [7], [8], [9], etc.

This paper suggests a new calibration estimator for population mean under stratified random sampling using a new calibration constraint which includes the median of the auxiliary variable. The use of median makes the estimator more efficient.

This paper is organized in six sections. Section I contains the introduction of the calibration approach and its related research work. Section II describes some notations defined for calibration approach and the estimators suggested by [7] and [9]. Section III proposes calibration estimator in stratified random sampling. Section IV considers extension of the results so obtained in case of double sampling when population median of auxiliary variable is not known. Section V deals with the simulation study using R-software to check the performance of the proposed estimators with the estimators suggested by [7] and [9]. Section VI concludes the research findings of the paper.

II. NOTATIONS DEFINED IN CALIBRATION APPROACH

Let us consider a heterogeneous population U of size N which has been divided into L homogeneous subgroups called strata consisting of N_h units in h^{th} stratum such that $\sum_{h=1}^{L} N_h = N$. A sample of size n_h is drawn using simple random sampling without replacement (SRSWOR) from the h^{th} stratum such that $\sum_{h=1}^{L} n_h = n$, where n is the required sample size. Suppose y_{hi} and x_{hi} is the i^{th} unit of the study and auxiliary variables, respectively, in the h^{th} stratum for i= 1, 2, ..., n_h and h = 1, 2, …, L. $W_h = \frac{N_h}{N}$ is the h^{th} stratum weight and $f_h = \frac{n_h}{N_h}$ is the h^{th} sample fraction.

The objective is to estimate the population parameter, say mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. The calibration estimator under the stratified random sampling for population mean $\bar{Y}$ defined by [10] is given as

$$\bar{y}_{tr} = \sum_{h=1}^{L} \Omega_h \bar{y}_h \tag{1}$$

where Ω_h ; h = 1, 2, …, L are the new calibrated weights obtained by minimizing the Chi-square distance measure $\sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{Q_h W_h}$, subject to the two calibration constraints:

$$\sum_{h=1}^{L}\Omega_h\bar{x}_h = \sum_{h=1}^{L}W_h\bar{X}_h \tag{2}$$

$$\sum_{h=1}^{L}\Omega_h s_{hx}^2 = \sum_{h=1}^{L}W_h S_{hx}^2 \tag{3}$$

where $\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$ and $\bar{X}_h = \frac{\sum_{i=1}^{N_h} X_{hi}}{N_h}$ are the h^{th} stratum sample and population means of the auxiliary variable, respectively.

$s_{hx}^2 = \sum_{i=1}^{n_h}\frac{(x_{hi}-\bar{x}_h)^2}{(n_h-1)}$ and $S_{hx}^2 = \sum_{i=1}^{N_h}\frac{(X_{hi}-\bar{X}_h)^2}{(N_h-1)}$ are the h^{th} stratum sample and population variances of the auxiliary variable, respectively.

Minimization of Chi-square distance measure, subject to the calibration constraints mentioned above, the calibrated estimator given by [10] is

$$\bar{y}_{tr} = \sum_{h=1}^{L}W_h\bar{y}_h + \hat{\beta}_{t1}\sum_{h=1}^{L}W_h(\bar{X}_h-\bar{x}_h) + \hat{\beta}_{t2}\sum_{h=1}^{L}W_h(S_{hx}^2-s_{hx}^2) \tag{4}$$

where

$$\hat{\beta}_{t1} = \frac{(\sum_{h=1}^{L}W_hQ_hs_{hx}^4)(\sum_{h=1}^{L}W_hQ_h\bar{x}_h\bar{y}_h) - (\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2)(\sum_{h=1}^{L}W_hQ_h\bar{y}_hs_{hx}^2)}{(\sum_{h=1}^{L}W_hQ_hs_{hx}^4)(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2) - (\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2)^2}$$

$$\hat{\beta}_{t2} = \frac{(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2)(\sum_{h=1}^{L}W_hQ_h\bar{y}_hs_{hx}^2) - (\sum_{h=1}^{L}W_hQ_hs_{hx}^2)(\sum_{h=1}^{L}W_hQ_h\bar{x}_h\bar{y}_h)}{(\sum_{h=1}^{L}W_hQ_hs_{hx}^4)(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2) - (\sum_{h=1}^{L}W_hQ_h\bar{x}_hs_{hx}^2)^2}$$

Similarly, the calibration estimator for population mean $\bar{Y}$ under stratified random sampling given by Nidhi et al. (2017) is

$$\bar{y}_{nc} = \sum_{h=1}^{L}\Omega_h\bar{y}_h \tag{5}$$

where Ω_h are the new calibrated weights obtained by minimizing the Chi-square distance measure $\sum_{h=1}^{L}\frac{(\Omega_h-W_h)^2}{Q_hW_h}$, subject to the two calibration constraints:

$$\sum_{h=1}^{L}\Omega_h\bar{x}_h = \sum_{h=1}^{L}W_h\bar{X}_h \tag{6}$$

$$\sum_{h=1}^{L}\Omega_h = 1 \tag{7}$$

Minimizing Chi-square distance measure, subject to the calibration constraints, the estimator given by [7] is

$$\bar{y}_{nc} = \sum_{h=1}^{L}W_h\bar{y}_h + \hat{\beta}_{nc}(\bar{X} - \sum_{h=1}^{L}W_h\bar{x}_h) \tag{8}$$

where

$$\hat{\beta}_{nc} = \frac{(\sum_{h=1}^{L}W_hQ_h)(\sum_{h=1}^{L}W_hQ_h\bar{x}_h\bar{y}_h) - (\sum_{h=1}^{L}W_hQ_h\bar{x}_h)(\sum_{h=1}^{L}W_hQ_h\bar{y}_h)}{(\sum_{h=1}^{L}W_hQ_h)(\sum_{h=1}^{L}W_hQ_h\bar{x}_h^2) - (\sum_{h=1}^{L}W_hQ_h\bar{x}_h)^2}$$

III. PROPOSED CALIBRATION ESTIMATOR

In this paper, we present a new calibration estimator in stratified random sampling using median, a positional average of the auxiliary variable in defining the calibration constraints. The proposed estimator is defined as

$$\bar{y}_{md} = \sum_{h=1}^{L}\Omega_h\bar{y}_h \tag{9}$$

where the calibration weights Ω_h ; (h = 1, 2, …, L) are chosen to minimize the Chi-square distance measure given as

$$\sum_{h=1}^{L}\frac{(\Omega_h-W_h)^2}{W_hQ_h} \tag{10}$$

subject to the following calibration constraint

$$\sum_{h=1}^{L}\Omega_h m_h = \sum_{h=1}^{L}W_h M_h \tag{11}$$

where m_h and M_h are the sample and population median of auxiliary variable, respectively.

The Lagrange function is defined as

$$L = \sum_{h=1}^{L}\frac{(\Omega_h-W_h)^2}{W_hQ_h} - 2\lambda(\sum_{h=1}^{L}\Omega_h m_h - \sum_{h=1}^{L}W_hM_h) \tag{12}$$

where λ is the Lagrange's multipliers. To determine the optimum value of Ω_h , we differentiate the Lagrange function given in equation (17) with respect to Ω_h and equate it to zero. Thus, the calibration weight can be obtained as

$$\Omega_h = W_h + \lambda(W_hQ_hm_h) \tag{13}$$

Here λ is determined by substituting the value of Ω_h from equation (13) to equation (11), so this leads to a calibrated weight given as

$$\Omega_h = W_h + W_hQ_hm_h\left[\frac{\sum_{h=1}^{L}W_h(M_h-m_h)}{\sum_{h=1}^{L}W_hQ_hm_h^2}\right] \tag{14}$$

After substituting the value of Ω_h from equation (14) to (9), we obtain the proposed calibrated estimator as

$$\bar{y}_{md} = \sum_{h=1}^{L}W_h\bar{y}_h + \hat{\beta}_{md}\left[\sum_{h=1}^{L}W_h(M_h-m_h)\right] \tag{15}$$

where $\hat{\beta}_{md} = \frac{\sum_{h=1}^{L}W_hQ_hm_h\bar{y}_h}{\sum_{h=1}^{L}W_hQ_hm_h^2}$

Now consider the different values of Q_h to obtain the various forms of calibration estimator as follows:

Case I: When $Q_h = 1$

The calibration estimator becomes

$$\bar{y}_{md1} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md1} \sum_{h=1}^{L} W_h (M_h - m_h)\right]$$

where $\hat{\beta}_{md1} = \dfrac{\sum_{h=1}^{L} W_h \bar{y}_h m_h}{\sum_{h=1}^{L} W_h m_h^2}$

Case II: When $Q_h = \dfrac{1}{\bar{x}_h}$

The calibration estimator becomes

$$\bar{y}_{md2} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md2} \sum_{h=1}^{L} W_h (M_h - m_h)\right]$$

where $\hat{\beta}_{md2} = \dfrac{\sum_{h=1}^{L} W_h \bar{y}_h \dfrac{m_h}{\bar{x}_h}}{\sum_{h=1}^{L} W_h \dfrac{m_h^2}{\bar{x}_h}}$

Case III: When $Q_h = \dfrac{1}{m_h}$

The calibration estimator becomes

$$\bar{y}_{md3} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md3} \sum_{h=1}^{L} W_h (M_h - m_h)\right]$$

where $\hat{\beta}_{md3} = \dfrac{\sum_{h=1}^{L} W_h \bar{y}_h}{\sum_{h=1}^{L} W_h m_h}$

IV. DOUBLE SAMPLING

The above result can also be extended in case of stratified double sampling when the value of median is not known for each stratum. In this technique, a preliminary sample of size n'_h units is drawn by SRSWOR as a first phase sample and a subsample of n_h units is drawn from the preliminary sample of size n'_h units by SRSWOR. Let $\bar{x}_h^* = \dfrac{1}{n'_h}\sum_{i=1}^{n'_h} x_{hi}$ is the first phase sample mean. $\bar{x}_h = \dfrac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ and $\bar{y}_h = \dfrac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ are the second phase sample means of auxiliary variable and study variable, respectively. Thus, the proposed calibration estimator in case of stratified double sampling is given as

$$\bar{y}_{md}^{d} = \sum_{h=1}^{L} \Omega_h^* \bar{y}_h \tag{16}$$

where the calibration weights Ω_h^*; (h = 1, 2, …, L) are chosen to minimize the Chi-square distance measure

$$\sum_{h=1}^{L} \frac{(\Omega_h^* - W_h)^2}{W_h Q_h} \tag{17}$$

subject to the following calibration constraint

$$\sum_{h=1}^{L} \Omega_h^* m_h = \sum_{h=1}^{L} W_h m_h^* \tag{18}$$

where m_h^* is the first phase sample median of the auxiliary variable.

The Lagrange function is given as

$$L = \sum_{h=1}^{L} \frac{(\Omega_h - W_h)^2}{W_h Q_h} + 2\lambda\left(\sum_{h=1}^{L} \Omega_h^* m_h - \sum_{h=1}^{L} W_h m_h^*\right) \tag{19}$$

where λ is the Lagrange's multiplier. For finding the optimum value of Ω_h^*, we differentiate the Lagrange function with respect to Ω_h^* and equate it to zero. Thus the calibration weight can be obtained as

$$\Omega_h^* = W_h + W_h Q_h m_h \left[\frac{\left(\sum_{h=1}^{L} W_h (m_h^* - m_h)\right)}{\left(\sum_{h=1}^{L} W_h Q_h m_h^2\right)}\right] \tag{20}$$

After substituting the value of Ω_h^* from equation (20) to (16), the proposed calibrated estimator can be obtained as

$$\bar{y}_{md}^{d} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md} \sum_{h=1}^{L} W_h (m_h^* - m_h)\right] \tag{21}$$

where

$$\hat{\beta}_{md} = \frac{\sum_{h=1}^{L} W_h Q_h m_h \bar{y}_h}{\sum_{h=1}^{L} W_h Q_h m_h^2}$$

Now the different forms of the calibration estimator for different values of Q_h are considered as

Case I: When $Q_h = 1$

The calibration estimator becomes

$$\bar{y}_{md1}^{d} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md1} \sum_{h=1}^{L} W_h (m_h^* - m_h)\right]$$

where $\hat{\beta}_{md1} = \dfrac{\sum_{h=1}^{L} W_h \bar{y}_h m_h}{\sum_{h=1}^{L} W_h m_h^2}$

Case II: When $Q_h = \dfrac{1}{\bar{x}_h}$

The calibration estimator becomes

$$\bar{y}_{md2}^{d} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md2} \sum_{h=1}^{L} W_h (m_h^* - m_h)\right]$$

where $\hat{\beta}_{md2} = \dfrac{\sum_{h=1}^{L} W_h \bar{y}_h \dfrac{m_h}{\bar{x}_h}}{\sum_{h=1}^{L} W_h \dfrac{m_h^2}{\bar{x}_h}}$

Case III: When $Q_h = \dfrac{1}{m_h}$

The calibration estimator becomes

$$\bar{y}^d_{md3} = \sum_{h=1}^{L} W_h \bar{y}_h + \left[\hat{\beta}_{md3} \sum_{h=1}^{L} W_h (m_h^* - m_h)\right]$$

where $\hat{\beta}_{md3} = \dfrac{\sum_{h=1}^{L} W_h \bar{y}_h}{\sum_{h=1}^{L} W_h m_h}$

V. Simulation Study

To compare the performance of the proposed estimators, real data of district wise wheat production of Uttar Pradesh for the year 2008-09 (www.agricoop.nic.in) are considered. In wheat production data, there are 63 districts, divided into 3 strata of varying sizes. A stratified random sample according to proportional allocation is drawn using SRSWOR. A simulation study generating 50,000 samples is done using R-software. For the comparison purpose, the empirical absolute relative bias and MSE, percentage relative efficiency of the estimators $\bar{y}_{tr}, \bar{y}_{nc}$ and $\bar{y}_{md}$ are calculated. Let $\bar{y}$ denotes an estimator of $\bar{Y}$ and $\bar{y}_k$ denotes the estimated value of $\bar{y}$ for the k^{th} sample (k = 1, 2, …, 50000). The empirical absolute relative bias of $\bar{y}$ can be calculated as

$$ARB(\bar{y}) = \frac{1}{50000} \sum_{k=1}^{50000} \left| \frac{\bar{y}_k - \bar{Y}}{\bar{Y}} \right|$$

The MSE of $\bar{y}$ is determined as

$$MSE(\bar{y}) = \frac{1}{50000} \sum_{k=1}^{50000} \left[\bar{y}_k - \bar{Y}\right]^2$$

The percent relative efficiency of the proposed estimator $\bar{y}_{md}$ with respect to the estimator $\bar{y}$ is computed as

$$\%RE = \frac{MSE(\bar{y})}{MSE(\bar{y}_{md})} \times 100$$

The results obtained from simulation study are given in Table 1 and 2 for stratified sampling and in Table 3 and 4 for stratified double sampling.

Table 1: Absolute Relative Bias (Stratified Sampling)

Q_h	$ARB(\bar{y}_{tr})$	$ARB(\bar{y}_{nc})$	$ARB(\bar{y}_{md})$
1	0.0021	0.0003	0.0156
$\frac{1}{\bar{x}_h}$	0.0006	0.0001	0.0153
$\frac{1}{m_h}$	0.0011	0.0003	0.0155

Table 2: MSE and %RE (Stratified Sampling)

Q_h		$\bar{y}_{tr}$	$\bar{y}_{nc}$	$\bar{y}_{md}$
1	MSE	1378961568.39	203457455.86	132438081.87
	%RE	100	677.76	1041.21
$\frac{1}{\bar{x}_h}$	MSE	1379114324.45	202374902.62	136496613.39
	%RE	100	681.47	1010.37
$\frac{1}{m_h}$	MSE	1449140610.60	203537650.21	139990467.32
	%RE	100	711.98	1035.17

Table 3: Absolute Relative Bias (Stratified Double Sampling)

Q_h	$ARB(\bar{y}^d_{tr})$	$ARB(\bar{y}^d_{nc})$	$ARB(\bar{y}^d_{md})$
1	0.0011	0.0004	0.0003
$\frac{1}{\bar{x}_h}$	0.0004	0.0002	0.0003
$\frac{1}{m_h}$	0.0010	0.0004	0.0003

Table 4: MSE and %RE (Stratified Double Sampling)

Q_h		$\bar{y}^d_{tr}$	$\bar{y}^d_{nc}$	$\bar{y}^d_{md}$
1	MSE	843727881.60	174662453.06	126191226.4850
	%RE	100	483.062	668.6106
$\frac{1}{\bar{x}_h}$	MSE	815126566.88	174079397.80	126380475.1870
	%RE	100	468.2499	644.9782
$\frac{1}{m_h}$	MSE	841363974.62	174610670.15	126205879.8837
	%RE	100	481.8514	666.6599

It is cleared from the above tables that the MSE of the proposed estimators are less than the estimators given by [7] and [10].

VI. Conclusion

The new calibration estimators proposed in this paper to estimate the population mean in case of stratified random sampling and stratified double sampling using median of the auxiliary variable in defining the calibration constraint are found to be more efficient than [7] and [10]. From the simulation study carried out on a real, it can be seen that the proposed estimators are having less MSE than the estimators given by [7] and [10]. As a result, the suggested estimators are more efficient for $Q_h = 1, \frac{1}{\bar{x}_h}$ and $\frac{1}{m_h}$ for the given dataset. Thus, it can be concluded that the proposed estimators are more efficient than the estimators given by [7] and [10].

REFERENCES

[1]. J.C. Deville, C.E. Sarndal, "*Calibration Estimators in Survey Sampling*", Journal of the American Statistical Association, Vol.**87**, No.**418**, pp.**376-382**, **1992**.

[2]. E.P. Clement, E.I. Enang, "*On the efficiency of ratio estimator over the regression estimator*", Communications in Statistics-Theory and Methods, Vol.**46**, No.**11**, pp.**5357-5367**, **2017**.

[3]. J-M Kim, E.A. Sungur, T-Y. Heo, "*Calibration approach estimators in stratified sampling*", Statistics and Probability Letters, Vol.**77**, pp.**99-103**, **2007**.

[4]. N. Koyuncu, C. Kadilar, "*Calibration estimator using different distance measures in stratified random sampling*", International Journal of Modern Engineering Research (IJMER), Vol.**3**, No.**1**, pp.**415-419**, **2013**.

[5]. N. Koyuncu, C. Kadilar, "*Calibration weighting in Stratified Random Sampling*", Communications in Statistics-Simulation and Computation, Vol.**45**, No.**7**, pp.**2267-2275**, **2016**.

[6]. A.M. Mouhamed, A.A. EI-sheikh, H.A. Mohamed, "*A new calibration estimator of Stratified random sampling*", Applied Mathematical Sciences, Vol.**9**, No.**35**, pp.**1735-1744**, **2015**.

[7]. Nidhi, B.V.S. Sisodia, S. Singh, S. K. Singh, "*Calibration approach estimation of the mean in stratified sampling and stratified double sampling*", Communications in Statistics-Theory and Methods, Vol.**46**, No.**10**, pp.**4932-4942**, **2017**.

[8]. S. Singh, R. Arnab, "*On Calibration of design weights*", METRON-International Journal of Statistics, Vol.**LXIX**, No.**2**, pp.**185-205**, **2011**.

[9]. D.S. Tracy, S. Singh, and R. Arnab, "*Note on Calibration in Stratified and Double Sampling*", Survey Methodology, Vol.**29**, No.**1**, pp.**99-104**, **2003**.

[10]. C.E. Sarndal, B. Swensson, J. Wretman, "*Model Assisted Survey Sampling*". Springer Verlag, **New York**, 1992.

Authors Profile

Dr. Neha Garg is currently working in School of Sciences, Indira Gandhi National Open University, New Delhi, India since 2010. Her research interests are in the areas of Statistics: Sampling Theory and Statistical Inference. She has 14 years of teaching and research experience. She has also published & presented several papers in refereed journals and conferences.

Ms. Menakshi Pachori is Pursuing Ph.D. (Statistics) from School of Sciences, Indira Gandhi National Open University, New Delhi, India. He is currently working in the area of sampling.