Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040313

A Chain Ratio-Type Estimator for Population Median under Double Sampling Design

Mursala Khan[1,*] *and Abdullah Yahia Al-Hossain*[2]

[1] Department of Mathematics, COMSATS Institute of Information Technology, Abbottabad, 22060, Pakistan
[2] Department of Mathematics, Faculty of Science, Jazan University, Jazan 2097, Saudi Arabia

Received: 28 Aug. 2015, Revised: 20 Oct. 2015, Accepted: 29 Oct. 2015
Published online: 1 Nov. 2015

Abstract: In this paper, we have proposed a chain ratio-type estimator for the population median under double sampling design using two auxiliary variables. The large sample properties of the proposed estimator are derived up to first order of approximation. For empirical study we have taken a real data sets from the literature of survey sampling, which shows that the proposed estimator has the higher percent relative efficiency over the [1] estimator, ratio, regression, [2], [3], [4] and [5] estimators.

Keywords: Median estimation, estimator, large sample properties, auxiliary variable, double sampling, mean square error, efficiency.

1 Introduction

The purpose of survey sampling is to gain statistical information about the finite population by selecting a probability sample from the finite population and measuring the required information about the units in the sample and estimating the finite population parameters such as means, median, totals, variance, and coefficient of variation etc. And in order to estimate the population parameters efficiently we incorporate auxiliary variables which is highly correlated with the study variable, availability of information on auxiliary variables in a finite population has become easily available from census data, previous surveys, administrative registers, experimentation and remote sensing data provide a wide and growing range of variables eligible to be employed to increase the precision of the estimators. Actually we have two types of distribution one is non-skewed distribution and the second one is skewed distribution and to estimate the population parameters for a non-skewed distribution arithmetic mean is the best measure of central tendency, while in estimating the population parameters for a skewed distribution median is the best measure of central tendency.

In the present paper, we are considering an estimator for population median under double sampling for a highly skewed distribution like the distribution of salaries, expenditure. Many statisticians and researchers have worked on the estimation of population median [1] and [6], [7], [8,9], [10] and [11].

Let us consider a finite population $U = \{U_1, U_2, U_3, \ldots, U_N\}$ of size N units, to estimate the population median M_y of the study variable$y$ taking values y_i , in the existence of two auxiliary variables say x and z taking values x_i and z_i respectively on the i^{th} population unit included in the sample of size n selected by simple random sampling without replacement. Let the population medians of the study variable and the auxiliary variables are M_y ,M_x and M_z respectively with corresponding second phase sample medians $\hat{M}_y$, $\hat{M}_x$ and $\hat{M}_z$ respectively, while $\hat{M}'_y$, $\hat{M}'_x$ and $\hat{M}'_z$ are the first phase sample medians respectively. Two phase sampling is used when the population median of the auxiliary variable x say M_x of closely related to the study variable is unknown, but information on the other economically auxiliary variable z closely correlated to x but compared to x remotely to y, is available for all the units in the finite population. In such a situation, we use two phase sampling. In this probability sampling scheme a large initial sample of size $n'(n' < N)$ is drawn from the population by using simple random sample without replacement sampling (SRSWOR) scheme and measure x and z to provide a good estimate for M_x . In the second phase, we draw a subsample of size n' from first phase sample of size n', i.e. $(n < n')$ by using SRSWOR or directly from the finite population Uand observed the study variable.

[*] Corresponding author e-mail: mursala.khan@yahoo.com

Suppose that the sample of size n observations of the study variable can be written in ascending order say $y_(1), y_(2), \ldots, y_(n)$. Further, suppose that t be an integer such that it satisfying the following inequality $y_t \geq M_y \leq y_{t+1}$ and let $p = \frac{t}{n}$ be the proportion of y values in the sample that are less than or equal to the median value M_y, an unknown population parameter. If denote the $Q_y(t)$ t-quantile of y then $\hat{M}_y = Q_y(0.5)$. [6] suggested the following matrices of proportions $p_{ij}(x,y)$, $p_{ij}(x,z)$ and $p_{ij}(y,z)$ respectively, and are given by

$X \backslash Y$	$Y \leq M_y$	$Y \geq M_y$	$Total$
$X \leq M_x$	$p_{11}(x,y)$	$p_{21}(x,y)$	$p_{.1}(x,y)$
$X \geq M_x$	$p_{12}(x,y)$	$p_{22}(x,y)$	$p_{.2}(x,y)$
$Total$	$p_{1.}(x,y)$	$p_{2.}(x,y)$	1

$X \backslash Z$	$Z \leq M_z$	$Z \geq M_z$	$Total$
$X \leq M_x$	$p_{11}(x,z)$	$p_{21}(x,z)$	$p_{.1}(x,z)$
$X \geq M_x$	$p_{12}(x,z)$	$p_{22}(x,z)$	$p_{.2}(x,z)$
$Total$	$p_{1.}(x,z)$	$p_{2.}(x,z)$	1

$Y \backslash Z$	$Z \leq M_z$	$Z \geq M_z$	$Total$
$Y \leq M_y$	$p_{11}(y,z)$	$p_{21}(y,z)$	$p_{.1}(y,z)$
$Y \geq M_y$	$p_{12}(y,z)$	$p_{22}(y,z)$	$p_{.2}(y,z)$
$Total$	$p_{1.}(y,z)$	$p_{2.}(y,z)$	1

Let the correlation coefficients between variables specified by the respective subscripts are given, as follows
$\rho_{xy} = 4[p_{11}(x,y) - 1]$, where $p_{11}(x,y)) = p(x \leq M_x \cap y \leq M_y)$,
$\rho_{yz} = 4[p_{11}(y,z) - 1]$, where $p_{11}(y,z)) = p(y \leq M_y \cap z \leq M_z)$ and
$\rho_{xz} = 4[p_{11}(x,z) - 1]$, where $p_{11}(x,z)) = p(x \leq M_x \cap z \leq M_z)$.

If $N \to \infty$, the distribution of the trivariate variables approaches a continuous distribution with marginal densities $f_y(y)$, $f_x(x)$ and $f_z(z)$ of y , x and z respectively. And let the probability density functions of y, x and z be $f_y(M_y)$, $f_x(M_x)$ and $f_z(M_z)$ respectively.

To obtain the large sample properties of the suggested estimator we use the following relative errors up to first order of approximation,and their expectation are given below
$e_0 = \frac{\hat{M}_y - M_y}{M_y}$, $e'_0 = \frac{\hat{M}_y{}' - M_y}{M_y}$, $e_1 = \frac{\hat{M}_x - M_x}{M_x}$, $e'_1 = \frac{\hat{M}_x{}' - M_x}{M_x}$, $e_2 = \frac{\hat{M}_z - M_z}{M_x}$ and $e'_1 = \frac{\hat{M}_z{}' - M_z}{M_z}$.
Such that $E(e_0) = E(e'_0) = E(e_1) = E(e'_1) = E(e_2) = E(e'_2) = 0$.
$E\left(e_0^2\right) = \theta\left[M_y f_y(M_y)\right]^{-2}$, $E\left(e_1^2\right) = \theta\left[M_x f_x(M_x)\right]^{-2}$, $E\left(e_2^2\right) = \theta\left[M_z f_z(M_z)\right]^{-2}$,
$E\left(e_0'^2\right) = E\left(e_0 e'_0\right) = \theta\left[M_y f_y(M_y)\right]^{-2}$, $E\left(e_1'^2\right) = E\left(e_1 e'_1\right) = \theta\left[M_x f_x(M_x)\right]^{-2}$, $E\left(e_2'^2\right) = E\left(e_2 e'_2\right) = \theta\left[M_z f_z(M_z)\right]^{-2}$,
$E(e_0 e_1) = \theta\rho_{yx}\left[M_y M_x f_y(M_y) f_x(M_x)\right]^{-1}$, $E(e_0 e_2) = \theta\rho_{yz}\left[M_y M_z f_y(M_y) f_z(M_z)\right]^{-1}$
$E(e_1 e_2) = \theta\rho_{xz}\left[M_x M_z f_x(M_x) f_z(M_z)\right]^{-1}$
$E(e_0 e'_1) = \theta_1\rho_{yx}\left[M_y M_x f_y(M_y) f_x(M_x)\right]^{-1}$, $E(e_0 e'_2) = \theta_1\rho_{yz}\left[M_y M_z f_y(M_y) f_z(M_z)\right]^{-1}$
$E(e_1 e'_2) = \theta_1\rho_{xz}\left[M_x M_z f_x(M_x) f_z(M_z)\right]^{-1}$.
Where $\theta = \frac{1}{n} - \frac{1}{N}$, $\theta_1 = \frac{1}{n_1} - \frac{1}{N}$ and $\theta_2 = \frac{1}{n} - \frac{1}{n_1}$.

The variance of the usual simple estimator $\hat{M}_g$ for population M_y median defined by [1] is given as
$V\left(\hat{M}_g\right) = \theta\left[2 f_y(M_y)\right]^{-2}$
The linear regression-type estimator $\hat{M}_{lr}$, for population median M_y , in two phase sampling scheme using information on two auxiliary variables is, given by

$$\hat{M}_{lr} = \hat{M}_y + \alpha\left(\hat{M}'_x - \hat{M}_x\right) + \beta\left(\hat{M}'_z - \hat{M}_z\right) \tag{1}$$

where α and β are unknown constants.
the minimum mean square error (MSE) up to first order approximation of the estimator $\hat{M}_{lr}$ is

$$MSE\left(\hat{M}_{lr}\right) = \theta\left[2 f_y(M_y)\right]^{-2}\left[\theta - \theta_2\rho_{yx}^2 - \theta_1\rho_{yz}^2\right] \tag{2}$$

and the optimum value of α and β are $\alpha = \frac{\rho_{yx} f_x(M_x)}{f_y(M_y)}$ and $\beta = \frac{\rho_{yz} f_z(M_z)}{f_y(M_y)}$

[2], defined a chain-ratio type estimator, the suggested estimator is, given by

$$\hat{M}_c = \hat{M}_y\left(\frac{\hat{M}'_x}{\hat{M}_x}\right)\left(\frac{M_z}{\hat{M}'_z}\right) \tag{3}$$

The MSE of the suggested estimator up to order first are, given by

$$MSE\left(\hat{M}_c\right)=\left\{2f_y\left(M_y\right)\right\}^{-2}\left[\theta+\theta_2\frac{M_y f_y\left(M_y\right)}{M_x f_x\left(M_x\right)}\left(\frac{M_y f_y\left(M_y\right)}{M_x f_x\left(M_x\right)}-2\rho_{yx}\right)+\theta_1\frac{M_y f_y\left(M_y\right)}{M_z f_z\left(M_z\right)}\left(\frac{M_y f_y\left(M_y\right)}{M_z f_z\left(M_z\right)}-2\rho_{yz}\right)\right] \quad (4)$$

[3], recommended the following power-chain ratio type estimator

$$\hat{M}_{sr}=\hat{M}_y\left(\frac{\hat{M}'_x}{\hat{M}_x}\right)^{\eta_1}\left(\frac{M_z}{\hat{M}'_z}\right)^{\eta_2} \quad (5)$$

where η_1 and η_2 are unknown constant, the minimum MSE of the estimator $\hat{M}_{sr}$ is given by

$$MSE\left(\hat{M}_{sr}\right)=\left\{2f_y\left(M_y\right)\right\}^{-2}\left[\theta-\theta_2\rho_{yx}^2-\theta_1\rho_{yz}^2\right] \quad (6)$$

and the optimum value of η_1 and η_2 are $\eta_1=\frac{\rho_{yx}M_x f_x(M_x)}{M_y f_y\left(M_y\right)}$ and $\eta_2=\frac{\rho_{yz}M_z f_z(M_z)}{M_y f_y\left(M_y\right)}$, respectively. [4], recommended a chain ratio estimator

$$\hat{M}_s=\hat{M}_y\left(\frac{\hat{M}'_x}{\hat{M}_x}\right)^{\alpha_1}\left(\frac{M_z}{\hat{M}'_z}\right)^{\alpha_2}\left(\frac{M_z}{\hat{M}_z}\right)^{\alpha_3} \quad (7)$$

where , α_1,α_2 and α_3 are the unknown constants,the first order minimum MSE and the optimum values of the constants are, given by

$$MSE\left(\hat{M}_s\right)=\left\{2f_y\left(M_y\right)\right\}^{-2}\left[\theta-\theta_1\rho_{yz}^2-\theta_2R_{y.xz}^2\right] \quad (8)$$

$\alpha_1=\frac{M_x f_x(M_x)\left(\rho_{xz}\rho_{yz}-\rho_{yx}\right)}{M_y f_y\left(M_y\right)\left(\rho_{xz}^2-1\right)}$, $\alpha_2=\frac{M_z f_z(M_z)\rho_{xz}\left(\rho_{xz}\rho_{yz}-\rho_{yx}\right)}{M_y f_y\left(M_y\right)\left(\rho_{xz}^2-1\right)}$ and $\alpha_3=\frac{M_z f_z(M_z)\left(\rho_{xz}\rho_{xy}-\rho_{yz}\right)}{M_y f_y\left(M_y\right)\left(\rho_{xz}^2-1\right)}$ respectively, also $R_{y.xz}^2=\frac{\rho_{xy}^2+\rho_{yz}^2-2\rho_{xy}\rho_{yz}\rho_{xz}}{1-\rho_{xz}^2}$ is the partial correlation coefficient among x, y and z. Using the known knowledge of range of the known variable z in addition with its population median, [5] recommended the following estimator for population median

$$\hat{M}_{gs}=\hat{M}_y\left(\frac{\hat{M}'_x}{\hat{M}_x}\right)^{\delta_1}\left(\frac{M_z+R_z}{\hat{M}'_z+R_z}\right)^{\delta_2}\left(\frac{M_z+R_z}{\hat{M}_z+R_z}\right)^{\delta_3} \quad (9)$$

where δ_1,δ_2 and δ_3 are the unknown constants, the minimum MSE of the estimator t_p is, given by

$$MSE\left(\hat{M}_{gs}\right)=\left\{2f_y\left(M_y\right)\right\}^{-2}\left[\theta-\theta_1\rho_{yz}^2-\theta_2R_{y.xz}^2\right] \quad (10)$$

where the optimum values of δ_1,δ_2 and δ_3 are $\delta_1=\frac{M_x f_x(M_x)\left(\rho_{xz}\rho_{yz}-\rho_{yx}\right)}{M_y f_y\left(M_y\right)\left(\rho_{xz}^2-1\right)}$,
$\delta_2=\frac{(M_z+R_z)f_z(M_z)\rho_{xz}\left(\rho_{xz}\rho_{yz}-\rho_{yx}\right)}{M_y f_y\left(M_y\right)\left(\rho_{xz}^2-1\right)}$ and $\delta_3=\frac{(M_z+R_z)f_z(M_z)\left(\rho_{xz}\rho_{xy}-\rho_{yz}\right)}{M_y f_y\left(M_y\right)\left(\rho_{xz}^2-1\right)}$ respectively.

2 The Proposed Estimator

Under the given sampling design scheme, wesuggesta chain ratio-type estimator for the population median using two auxiliary variables, the suggested estimator is given,as follows

$$\hat{M}_m=\hat{M}_y\left(\frac{\hat{M}_x}{\hat{M}'_x}\right)^{k_1}\left(\frac{M_z}{\hat{M}'_z}\right)^{k_2}+k_3\left(\hat{M}'_y-\hat{M}_y\right) \quad (11)$$

where k_1 , k_2 and k_3 are the unknown constants, whose values are to be found such that it makes the MSE of the suggested estimator minimum,using first order approximation technique, the MSE of the proposed estimator is minimum for the values of k_1 , k_2 and k_3, that is ,
$k_1=\frac{-2\rho_{xy}M_x f_x(M_x)}{\left(1-\rho_{xy}^2\right)M_y f_y\left(M_y\right)}$, $k_2=\frac{\rho_{yz}M_z f_z(M_z)}{M_y f_y\left(M_y\right)}$ and $k_3=\frac{1+\rho_{xy}^2}{1-\rho_{xy}^2}$ respectively, and the minimum MSE is, given by

$$MSE\left(\hat{M}_m\right)=\left\{2f_y\left(M_y\right)\right\}^{-2}\left[\theta-\theta_1\rho_{yz}^2-\frac{\theta_2\left\{1+4\rho_{xy}^2\left(1-\rho_{xy}^2\right)\right\}}{\left(1-\rho_{xy}^2\right)^2}\right] \quad (12)$$

3 Comparison

The proposed estimator will performed better than the other discussed existing estimators if the following conditions are fulfilled. From equation (1.1) and (2.2), we have $MSE\left(\hat{M}\right)_m < MSE\left(\hat{M}_g\right)$, if

$$\left[\theta_1\rho_{yz}^2 + \frac{\theta_2\left\{1+4\rho_{xy}^2\left(1-\rho_{xy}^2\right)\right\}}{\left(1-\rho_{xy}^2\right)^2}\right] > 0.$$

From equation (1.3) and (2.2), we have $MSE\left(\hat{M}\right)_m < MSE\left(\hat{M}_{lr}\right)$, if

$$\left[1+\rho_{xy}^2\left(1-\rho_{xy}^2\right)\left(3+\rho_{yx}^2\right)\right] > 0.$$

From equation (1.5) and (2.2), we have $MSE\left(\hat{M}\right)_m < MSE\left(\hat{M}_c\right)$, if

$$\left[\theta_1\left\{\rho_{yz}^2 + \frac{M_y f_y\left(M_y\right)}{M_z f_z\left(M_z\right)}\left(\frac{M_y f_y\left(M_y\right)}{M_z f_z\left(M_z\right)} - 2\rho_{yz}\right)\right\} + \theta_2\left\{\frac{1+4\rho_{xy}^2\left(1-\rho_{xy}^2\right)}{\left(1-\rho_{xy}^2\right)^2} + \frac{M_y f_y\left(M_y\right)}{M_x f_x\left(M_x\right)}\left(\frac{M_y f_y\left(M_y\right)}{M_x f_x\left(M_x\right)} - 2\rho_{yx}\right)\right\}\right] > 0.$$

From equation (1.7) and (2.2), we have $MSE\left(\hat{M}\right)_m < MSE\left(\hat{M}_{sr}\right)$, if

$$\left[1+\rho_{xy}^2\left(1-\rho_{xy}^2\right)\left(3+\rho_{yx}^2\right)\right] > 0.$$

From equation (1.9) and (2.2), we have $MSE\left(\hat{M}\right)_m < MSE\left(\hat{M}_s\right)$, if

$$\left[\frac{1+4\rho_{xy}^2\left(1-\rho_{xy}^2\right)}{\left(1-\rho_{xy}^2\right)^2} - R_{y.xz}^2\right] > 0.$$

From equation (1.11) and (2.2), we have $MSE\left(\hat{M}\right)_m < MSE\left(\hat{M}_{gs}\right)$, if

$$\left[\frac{1+4\rho_{xy}^2\left(1-\rho_{xy}^2\right)}{\left(1-\rho_{xy}^2\right)^2} - R_{y.xz}^2\right] > 0.$$

4 Empirical Study and Results

To show the performance of the proposed estimator numerically, we have considered two real data sets from the literature of survey sampling. The first population is taken from [12], while the second population is taken from MFA [13]. The description and the necessary data statistics of the populations are, given by

Population-I. [12].
y: the number of fish caught by marine recreational fisherman in 1995,
x: the number of fish caught by marine recreational fisherman in 1994,
z: the number of fish caught by marine recreational fisherman in 1993,
z: the number of fish caught by marine recreational fisherman in 1993.
$N = 69, n' = 24, n = 17, M_y = 2068, M_x = 2011, M_z = 2307, f_y\left(M_y\right) = 0.00014, f_x\left(M_x\right) = 0.00014,$
$f_z\left(M_z\right) = 0.00013, \rho_{yx} = 0.1505, \rho_{yz} = 0.3166, \rho_{xz} = 0.01431.$

Population-II. [13].
y: District-wise production of tomato (tons) in Pakistan in year 2003,
x: District-wise production of tomato (tons) in Pakistan in year 2002,
z: District-wise production of tomato (tons) in Pakistan in year 2001.
$N = 97, n' = 46, n = 33, M_y = 1242, M_x = 1233, M_z = 1207, f_y\left(M_y\right) = 0.00021, f_x\left(M_x\right) = 0.00022,$
$f_z\left(M_z\right) = 0.00023, \rho_{yx} = 0.2096, \rho_{yz} = 0.1233, \rho_{xz} = 0.01496.$

The Percent relative efficiencies (PRE) of the proposed and the existing estimators with respect to $\hat{M}_g$ are shown in Table 1.

Table 1: Percent relative efficiency

$Estimator$	$Population-I$	$Population-II$
$\hat{M}_g$	100.00	100.00
$\hat{M}_{lr}$	107.50	102.8
$\hat{M}_c$	67.29	62.68
$\hat{M}_{sr}$	107.55	102.83
$\hat{M}_s$	105.84	103.7
$\hat{M}_{gs}$	105.84	103.7
$\hat{M}_m$	201.36	227.27

5 Results and Conclusion

The results based on the above real data sets are given in Table 1, which clearly indicates that the proposed estimator $\hat{M}_m$ has the higher percent relative efficiency than the [1] estimator, ratio, regression, [2], [3],[4] and [5] estimators.

Acknowledgements

The authors are thankful to the Editor and to the learned referees for their valuable suggestions regarding the improvement of the paper.

References

[1] S. T. Gross. Median estimation in sample surveys, Proceedings of the Survey Research Methods Section of the American Statistical Association, 181-184, 1980.
[2] L. Chand. Some ratio-estimators based on two or more auxiliary variables, PhD dissertation, Iowa State University, Ames, Iowa, unpublished, 1975.
[3] S. K. Srivastava, S. Rani, B. B. Khare and S. R. Srivastava. A generalized chain ratio estimator for mean of finite population, Journal of Indian Society of Agricultural Statistics, **42**(1):108-117, 1990.
[4] S. Singh, H. P. Singh and L. N. Upadhyaya.Chain ratio and regression type estimators for median estimation in survey sampling,Statistical Papers, **48**(1): 23 - 46, 2006.
[5] S. Gupta, J. Shabbir and S. Ahmad. Estimation of median in two phase sampling using two auxiliary variable,Communication in Statistics-Theory and Methods, **37**(11): 1815 - 1822, 2008.
[6] A. Y. C. Kuk, and T. K. Mak. Median estimation in the presence of auxiliary information, Journal of Royal Statistical Society,B, **51**:261 - 269, 1989.
[7] M. R. Garcia and A. A. Cebrian. On estimating the median from survey data usingmultiple auxiliary information,Metrika,**54**: 59 - 76, 2001.
[8] S. Singh, A. H. Joarderand D. S. Tracy. Median estimation using double sampling,Australian and New Zealand Journal of Statistics, **43**:33 - 46, 2001.
[9] H. P. Singh, S. Singh and A. H. Joarder. Estimation of population median when mode of an auxiliary variable is known, Journal of Statistical Research, **37**(1): 57 - 63, 2003.
[10] S. Al and H. Cingi. New estimators for the population median in simple random sampling, Tenth Islamic Countries Conference on Statistical Sciences, held in New Cairo, Egypt, 2009.
[11] H. P. Singh and R. S. Solanki. Some Classes of estimators for the Population Median Using Auxiliary Information.Communications in Statistics-Theory and Methods, **42**:4222-4238, 2013.
[12] S. Singh. “Advances Sampling Theory and Applications.How Michael ’Selected’ Army”, Volume I and II, Kluwer Acadamic Publishers, The Netherlands, 2003.
[13] MFA. Crops Area Production, Ministry of Food, Agriculture and Livestock, Economic Wing, Islamabad, Pakistan, 2004.