ON ESTIMATION OF POPULATION MEAN USING KNOWN AUXILIARY VARIABLE

Summary of

Dissertation Submitted to

Babasaheb Bhimrao Ambedkar University
Established 1996
Accredited 'A' Grade by NAAC in 2015

(DEPARTMENT OF STATISTICS)

School of Physical and Decision Sciences

For the Degree of

MASTER OF PHILOSOPHY

In

STATISTICS

By

Shanya Baghel, MSc.

Enrollment No. 1199/16

Under the supervision of

Dr. Subhash Kumar Yadav

Assistant Professor

Department of Statistics

BBAU, Lucknow

SEPTEMBER 2019

SUMMARY

Chapter 1 consists of brief introduction of sampling theory and the importance of auxiliary information for the estimation of population parameters. The various existing estimators for the population mean estimation developed by researchers have also been reviewed. The summary of first chapter is given in the following paragraphs.

Information about something can be obtained either as a whole where each and every detail is considered or as a summary where essence of overall details is provided. In definitive terms former is considered as the Complete Enumeration and later as the Sampling. Sampling is inevitable part of human life such as taking a handful of rice to check the quality of rice in the sack, meeting one or two person of any country to tell the nature of the people of that country, blood group testing etc.

Sampling is not just a partial coverage of whole; it is the Science of controlling and measuring reliability of useful Statistical information through the theory of Probability. Sample being part of the whole, its extrapolation leads to error called as Sampling Error. The aim of statistician is to reduce the sampling errors either by formulating suitable sampling schemes or by developing efficient estimators of the parameters or both. We have developed the efficient estimators for the population parameters in our dissertation. The simplest method of sampling is the Simple Random Sampling (SRS) Scheme where each unit of the population at the each draw has equal probability of being selected in the sample. SRS Scheme can be with replacement (WR) or without replacement (WOR). We have considered the SRSWOR Scheme in our dissertation.

Sampling is the cost effective method of drawing the valid inferences about the population parameters. These Population parameters are mean, variance, skewness, kurtosis etc. We have studied the estimation of population mean. If we denote the characteristic under study as y, simple random sample mean $\bar{y}$ is the basic estimator of the population mean. The auxiliary information which itself is independent but the characteristic under study has some degree of relationship with it; is used for improved estimation of parameters to increase the efficiencies of the estimators. For example to analyze the annual household expenditure we can take household size as auxiliary information.

The use of auxiliary information dates back to year 1926(Bowley), Newman (1934, 1938) who dealt with the stratification of finite population and putting forward a theoretical criticism of non-random (purposive) sampling. Watson (1937) gave the regression method of estimation in which there is reasonable degree of linear relationship between study variable (say y) and auxiliary variable (say x). Cochran (1940, 1942) propounded the ratio method of estimation in which relationship between auxiliary variable and study variable is of rough proportionality. Hansen and Hurwitz (1943) were the first to suggest use of auxiliary information in the varying probability sampling scheme.

Ratio estimators are the one of the most widely used estimators for population mean or total when the auxiliary and study variable are positively correlated. When correlation is

negative product estimator is used. Though the precision of regression estimator is higher than the ratio estimator yet for the large scale surveys ratio estimator would be the wiser choice because of its simplicity. Regression estimator is preferred when there is moderate or lower correlation or some kind of linear relationship between auxiliary variable and study variable.

Auxiliary information can be used at designing stage, sampling stage or estimation stage. The use of the auxiliary information at estimation stage is dealt in our present study. Several estimators have been developed till date to increase the efficiency using auxiliary variable. Parameters of the auxiliary variables such as mean, median, coefficient of variation etc have been used to increase the efficiency of the estimators. Many of such estimators are considered in the literature.

Goodman and Hartley (1958) were concerned with the modification of ratio estimator which would lead to an unbiased estimator. They considered both situations where sample size in each stratum was small and when large. Walsh (1970) modified the denominator of the ratio estimator and proposed the generalized form of the estimate of population total. Chakrabarty (1979) presented some ratio estimators. Sahai (1979) provided an efficient variant of the product and ratio estimator.

Ray, Sahai and Sahai (1979) suggested ratio and product type transformed estimators obtained through parametric linear combination of mean per unit estimator and ratio estimator; mean per unit estimator and product estimator respectively. Sahai and Ray (1980) proposed two-parameter families of ratio-type and product-type estimators for a finite population mean based on simple random samples of observations on the variable of interest and a concomitant variable. Using some prior information they showed that the families contain estimators which have in practical situations lower mean squared error than the usual ratio, product and sample mean estimators.

Sisodia and Dwivedi (1981) modified the ratio estimator using the coefficient of variation of auxiliary variable with increased efficiency. Srivastava (1983) provided the predictive estimation of finite population mean using the product estimator. Chauby, Singh and Dwivedi (1984) wrote down a note which points out the derivation of regression estimator through an optimality consideration over a class of estimators generating a Generalised Product and dual to ratio estimators. Bahl and Tuteja (1991) introduced new ratio and product type exponential estimators for estimating the mean of the finite population using information on single auxiliary variable. Upadhyaya and Singh (1999) by combining the coefficient of kurtosis and coefficient of variation of auxiliary variable proposed the estimator of the finite population mean with a greater precision. They also obtained the unbiased version of their suggested estimators using interpenetrating subsample design and Jack-knife technique.

Kadilar and Cingi (2003) studied on the chain ratio type estimators. Singh and Tailor (2003) used the known correlation coefficient for efficiently estimating the population mean. Singh, Tailor and Kakran (2004) improved the estimator of population mean using power transformation. Al-Omari, Jemain and Ibrahim (2009) suggested modified ratio estimators of the population mean of the variable of interest involving the first or third quartiles of an auxiliary variable that is correlated with the variable of interest. They also investigated the newly suggested estimators under simple random sampling (SRS) and ranked set sampling (RSS) methods. Yan and Tian (2010) obtained the ratio method to the mean estimation using coefficient of skewness of auxiliary variable.

Subramani and Kumarpandiyan (2012) modified the ratio estimators using function of quartiles, median and coefficient of skewness of auxiliary variable. Jeelani, Maqbool and Mir (2013) transformed the ratio estimators using linear combination of coefficient of skewness and quartile deviation. Singh, Solanki and Singh (2016) suggested the ratio-type and product-type exponential estimators of the population mean of a study variable through predictive approach using Bahl and Tuteja (1991) ratio-type and product-type exponential estimators as a predictor of the mean of the unobserved units of the population. Properties of the suggested estimators were also studied up to first order of approximation in simple random sampling using information on an auxiliary variable. Swain (2014) suggested an alternative ratio type exponential estimator with increased efficiency.

Singh and Pal (2015) advocated the problem of estimating the finite population mean using auxiliary information in sample surveys. They suggested a new chain ratio-ratio-type exponential estimator and its properties were studied up to first degree of approximation. Generalized version of the suggested chain ratio-ratio-type estimator was also given along with its properties. Yadav and Mishra (2015) developed an improved estimator using predictive method of estimation utilizing auxiliary information. Jerajuddin and Kishun (2016) improved the ratio estimator using size of the sample. Kadilar (2016) developed a new exponential type estimator which came out to be more efficient than the existing ones. Singh and Solanki (2016) provided a class of new estimators so as to gain the higher precision. Subramani (2016) utilized the median of study variable to obtain the much better estimator of the population mean.

Vishwakarma, Singh, Gupta, Pareek (2016) used the combination of determined constants to obtain the efficient estimator for the population mean. Soponviwatkul and Lawson (2017) derived new ratio estimators for estimating population mean in SRS using a coefficient of variation, correlation coefficient and a regression coefficient. Yadav and Dixit (2019) developed a class of ratio estimators combining various parameters of auxiliary variables and used it for the estimation of average yield of peppermint.

In practice almost all surveys suffer from non-response. The problem of non-response often happens due to the refusal of the subject, absenteeism and sometimes due to the lack of information. The pioneering work of Hansen and Hurwitz (1946), assumed that a sub sample of initial non-respondents is re-contacted with a more expensive method, suggesting the first attempt by mail questionnaire and the second attempt by a personal interview. In estimating population parameters such as the mean, total or ratio, sample survey experts sometimes use auxiliary information to improve precision of the estimates.

Rao (1986) proposed an improved ratio estimator with sub sampling the non respondent. Khare and Sinha (2007) suggested the estimation of the ratio of the two population means (response class and non response class) using multi auxiliary characteristics in the presence of non-response. Singh and Kumar (2008) provided a regression approach to the estimation of finite population mean in presence of non response. Singh and Kumar (2009) derived a general procedure for estimating the population mean in presence of non response under double sampling using auxiliary information. Kumar and Bhougal (2011) proposed a modified ratio-product type exponential estimator to estimate the finite population mean of the study variable in presence of non-response in different situations viz. (i) population mean of auxiliary variable is known, and (ii) population mean of auxiliary variable is unknown. The expressions of biases and mean squared error of the proposed estimators had been obtained under large sample approximation using single as well as double sampling. Khare and Sinha (2012) considered the problem of estimation of ratio of two population means using multivariate auxiliary characters with known population means under incomplete information. They proposed the general class of estimators and studied its properties. Chanu (2015) suggested an improved exponential ratio cum exponential dual ratio estimator of finite population mean in presence of non response. Pal and Singh (2016) derived a finite population mean estimation through a two parameter ratio estimator using auxiliary information in presence of non response. Zubir et al. (2018) proposed an efficient exponential estimator for population mean of study variable, with two auxiliary variables using two phase sampling scheme with sub sampling techniques in presence of non-response. The two situations with known population means of auxiliary variables, incomplete information on study variables and incomplete/complete information on auxiliary variables had been considered. Expressions for bias and mean square error had been derived. They proved their estimator to be considerably efficient.

The present thesis "On Estimation of Population Mean Using Known Auxiliary Variable" is confined in the use of auxiliary variable at the estimation stage. Inspired by many authors we have proposed estimators in different situations. The sampling scheme is taken as to be the Simple Random Sampling without replacement. Our interest lies in developing the ratio type of estimators because they are simpler to implement than the regression estimator. The data we used in the chapters to justify results are secondary data.

Chapter 2 consists of an improved class of estimators for population mean of study variable using the informations related to an auxiliary variable combined with the median of the study variable under the Simple Random Sampling Scheme. We have shown that many of the existing estimators of population mean are the members of our proposed class. The Bias and MSE of our proposed class is derived upto the first order of approximation. Minimum values of Bias and MSE is obtained by optimizing the chracterizing scalar. Bias and MSE has also been compared with the existing estimators. Finally we have suggested some new members of our proposed class which are more efficient than the existing ones numerically.

Chapter 3 provides a remedy for improved estimation of population mean of study variable, using the information related to two auxiliary variables in the situation of non response under Simple Random Sampling Scheme. The Bias and MSE of the proposed estimator are derived upto the first order of approximation. The minimum value of the MSE is also obtained by optimizing the characterizing scalers. The Bias and MSE have also been compared with the considered existing competing estimators both theoretically and empirically. The theoretical conditions are verified for the proposed estimator to be more efficient than the considered existing estimators using the two natural populations.

Chapter 4 takes the situation of developing a class of efficient estimator without the use of median of study variable. Bias and MSE are found and a comparative study with the conventional mean per unit unbiased estimator and existing competing estimators is provided. Empirical study showing the increased efficiency of proposed class of estimator over the mean per unit estimator and many existing estimators are also included as an illustration.

Chapter 5 is an improvement over the estimator of chapter 4. It provides a very useful class of estimators for estimation of population mean. The constants mentioned in the proposed class of estimator are chosen such that the mean squared error is minimum. By measuring the optimum value of the major constant we obtain an unbiased estimator with minimum MSE. Results are compared theoretically as well as empirically.

References

Al-Omari, A.I., Jemain, A.A.,and Ibrahim, K. (2009): New Ratio Estimators of the Mean using Simple Random Sampling and Ranked set Sampling Methods. *Investigacion Operacional*, 30(2), 97-108.

Bahl, S. and Tuteja, R.K. (1991) : Ratio and product type exponential estimator.*Information and Optimization Sciences* , XII (I), 159-163.

Chakrabarty, R.P. (1979) : Some ratio estimators.*Journal of the Indian Society of Agricultural Statistics*, 31(1), 49–57.

Chanu, W. W. and Singh B. K. (2015) : Improved Exponential Ratio cum Exponential Dual to Ratio Estimator of Finite Population Mean in Presence of Non-Response. *J. Stat. Appl. Pro.*, 4(1), 103-111.

Chauby, Y.P., Singh, M., Dwivedi, T.D. (1984) : A note on an optimality of the regression estimator.*Biometrical J.*, 26(4), 465-467.

Cochran, W.G. (1940): The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *The Journal of Agric. Science*, 30, 262-275.

Goodman, L.A. (1960) : On the exact variance of products *J. Amer. Stat. Ass.* , 55, 708-713.

Goodman, L. A. and Hartley, H.O. (1958): The Precision of Unbiased Ratio-Type Estimators. *Journal of the American Statistical Association*, 53, 282, 491-508.

Hansan, M.H., Hurwitz, W.N. (1946) : The problem of non-response in sampling surveys. *Journal of American Statistical Association*, 41 , 517-529.

H. Pandey, S.K. Yadav and A.K. Shukla (2011) : An Improved General Class of Estimators, Estimating Population Mean using Auxiliary Information.*International Journal of Statistics and Systems* ,Vol.6, 1, 1–7.

Jeelani, M.I., Maqbool, S., and Mir, S.A. (2013): Modified Ratio Estimators of Population Mean using Linear Combination of Coefficient of Skewness and Quartile Deviation. *International Journal of Model Mathematical Sciences*, 6(3), 174-183.

Jerajuddin, M. and Kishun, J. (2016) : Modified Ratio Estimators for Population Mean Using Size of the Sample, Selected From Population.*IJSRSET* , 2:2,10-16.

Kadilar, C., and Cingi, H. (2003) : A study on the chain ratio type estimator.*Hacettepe Journal of Mathematics and Statistics*, Vol.32, 105-108.

Kadilar, G.O. (2016) : A New Exponential Type Estimator for the Population Mean in Simple Random Sampling. *Journal of Modern Applied Statistical Methods* , 15:2 ,2017-2014 .

KHARE, B. B., SINHA, R, R. (2007) : Estimation of the ratio of the two population means using multi auxiliary characteristics in the presence of non-response. *In Statistical Techniques in Life Testing, Reliability, Sampling Theory and Quality Control.*, 1, 63-171

Khare, B. B., Sinha, R, R. (2012) : Improved classes of estimators for ratio of two means with with double sampling the non-respondents. *Statistica* , 49(3), 75-83.

Kumar, S., Bhougal, S. (2011) : Estimation of the population mean in presence of non-response. *Communications of the Korean Statistical Society*, 18(4), 537-548.

Muhammad Zubair, Asad Ali, Wajiha Nasir and Muhammad Rashad (2018) : An Exponential Estimator in presence of Non Response. 16^{th} *International Conference on Statistical Sciences*, 32, 209-222.

Pal, S. K. and H. P. Singh, H. P. (2016) : Finite Population Mean Estimation through a Two-Parameter Ratio Estimator Using Auxiliary Information in Presence of Non-Response. *JAMSI*, 12(2), 5-39.

Rao, P.S.R.S. (1986) : Ratio estimation with sub sampling the non-respondents. *Survey Methodology*, 12(2), 217–230.

Ray, S.K., Sahai, A., Sahai, A. (1979) : A note on ratio and product type estimators.*Ann. Inst. Statist. Math.*, 31(1), 141-144.

Reddy, V.N. (1974) : On a transformed ratio method of estimation.*Sankhya C* , 36(1), 59-70.

SAHAI, A. (1979) : An efficient variant of the product and ratio estimator.*Statistica Neerlandica*, 33(1),27-35.

SAHAI, A. AND RAY, S.K. (1980) : Efficient families of ratio and product type estimators.*Biometrika*, 67(1), 211–215.

SINGH, D., CHAUDHARY, F.S. (1986) : Sample Survey Designs. *New Age Int.(P) Ltd.*.

SINGH, H.P., KUMAR, S. (2008) : A regression approach to the estimation of the finite population mean in the presence of non-response. *Aust. N. Z. J. Stat.*, 50(4), 395–408.

SINGH, H.P., KUMAR, S. (2009) : A general procedure for estimating the population mean in presence of non-response under double sampling using auxiliary information. *SORT*, 33(1), 71-83.

SINGH, PAL AND SOLANKI (2016) : A new class of estimators of finite population mean in sample surveys *Communications in Statistics-Theory and Methods* , 46:6, 2630-2637 .

SRIVASTAVA, S.K. (1967) : An estimator using auxiuary information in sample surveys.*GkuftU Sfatist.Ass.Bull.*, 16,121-132.

SINGH, H.P. AND PAL, S.K. (2015) :A New Chain Ratio-Ratio-Type Exponential Estimator Using Auxiliary Information in Sample Surveys.*International Journal of Mathematics and its Applications*,3 ,4:B, 37-46.

SINGH, H. P., SOLANKI, R. S., SINGH, A. K. (2014) : Predictive Estimation of Finite Population Mean Using Exponential Estimators.*Statistika: Statistics and Economy Journal*, 94 (1), 41–53.

SINGH, H.P., AND TAILOR, R. (2003): Use of known Correlation Coefficient in Estimating the Finite Population Mean. *Statistics in Transition*, 6(4), 550-560.

SINGH, H.P., TAILOR, R., AND KAKRAN, M.S. (2004): An improved Estimator of population mean using power transformation. *Journal of Indian Society of Agricultural Statistics*, 58(2), 223-230.

SISODIA, B.V.S. AND DWIVEDI, V.K (1981) : A modified ratio estimator using coefficient of variation of auxiliary variable., Journal of the Indian Society of Agricultural Statistics, 33, 13-18.

SOPONVIWATKUL, K AND LAWSON, N. (2017) : New ratio estimators for estimating population mean in simple random sampling using a coefficient of variation, correlation coefficient and a regression coefficient. *Gazi University Journal of Science* , 30(4) , 610-621.

SRIVASTAVA, S.K. (1983) : Predictive estimation of finite population mean using product estimator.*Metrica*, 30(1), 93-99.

SUBRAMANI (2016) : A new median based ratio estimator for estimation of the finite population mean. *Statistics in Transition new series* ,17:4 ,591-604.

SUBRAMANI, J. AND KUMARAPANDIYAN, G. (2012a): Modified Ratio Estimators for Population Mean using Function of Quartiles of Auxiliary Variable. *Bonfring International Journal of Industrial Engineering and Management Science,* 2(2), pp. 19-23,2012.

SUBRAMANI, J. AND KUMARAPANDIYAN, G. (2012b): Estimation of Population Mean using known Median and Coefficient of Skewness. *American Journal of Mathematics and Statistics*, 2(5), pp. 101-107.

SWAIN, A.K.P.C (2014) : On an improved ratio type estimator of finite population mean in sample surveys.*Revista Investigacion Operacional*, 35, 1, 49-57.

UPADHYAYA, L.N. AND SINGH, H.P. (1999): Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean. *Biometrical Journal,* 41, pp. 627-636.

VISHWAKARMA, SINGH, GUPTA, PAREEK (2016) : Improved ratio and product type estimators of finite population mean in simple random sampling. *Revista Investigacion Operacional* , 36:3, 70-76.

WALSH, J.E. (1970) : Generalization of ratio estimator for population total.*Sankhya A*, 32(1), 99-106.

WATSON, D.J. (1937) : The Estimation of leaf areas.*Jour. Agri. Sci.*,27, 474.

YADAV, S.K. AND MISHRA, S.S. (2015) : Developing Improved Predictive Estimator for Finite Population Mean Using Auxiliary Information.*Statistika*, 95:1, 76-85.

YADAV, S.K., DIXIT, M.K., DUNGANA, H.N., MISHRA, S.S. (2019): Improved Estimators for Estimating Average Yield Using Auxiliary Variable. *International Journal of Mathematical Engineering and Management Sciences*, Vol.x,No.x,xxx-xxx,20xx

YAN, Z. AND TIAN, B. (2010): Ratio Method to the Mean Estimation using Coefficient of Skewness of Auxiliary Variable. *ICICA 2010, Part II, CCIS*, 106, pp. 103-110.