SOME CONTRIBUTION TO THE THEORY OF NON-RESPONSE

ABSTRACT
of
THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

(A CENTRAL UNIVERSITY)

LUCKNOW

BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY

•LUCKNOW•

प्रज्ञा शील करूणा

ESTABLISHED 1996

FOR THE DEGREE OF

Doctor of Philosophy

IN

APPLIED STATISTICS

Submitted by

ABHAY PRATAP PANDEY

Supervisor

Dr. SURINDER KUMAR

Co-Supervisor

Dr. SHASHI BHUSHAN

DEPARTMENT OF APPLIED STATISTICS
SCHOOL FOR PHYSICAL SCIENCES
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
VIDYA VIHAR, RAEBARELI ROAD
LUCKNOW-226 025

ENROLMENT NUMBER: 395/11

2017

Abstract

The sampling theory mainly deals with problem of drawing the inference about the characteristics of a population based on a representative sample. While proposing a sampling strategy for the given problem at hand we use different sampling schemes along with various estimation methods for studying the characteristics of interest. In estimation of population characteristics ancillary information closely related to the main characteristic plays a very important role. In estimation of population characteristics, the population parameters can be estimated more accurately by making use of information on an auxiliary variable that is correlated with study variable. The ratio and regression methods of estimation are good examples in this context. In sampling theory, it is well known that the efficiency of the estimators of unknown population parameters of the study variable y can be increased by suitably using known information on an auxiliary variable x, which may be positively or negatively correlated with y.

In sampling theory mainly two types of errors are committed, one is sampling error and other is non-sampling error. The sampling error can be minimized by increasing the size of the sample but non sampling errors pose more severe problem in estimation part of sampling theory. The non-sampling error can occur at any one or more stages of the survey. These errors are broadly classified as non-response error, response error, tabulation error etc. In presence of non-response, the estimate of parameter of interest under conventional sampling strategies lead to serious bias as some units fail to respond thereby not covering the targeted population as per the sampling strategy.

The problem of non response was first studied by Hansen and Hurwitz (1946) related to a mailed questionnaire. Hansen and Hurwitz suggested a technique for adjustment of bias due to non-response. The procedure suggested by Hansen and Hurwitz results in an unbiased estimator of population mean. Hansen and Hurwitz suggested a technique for dealing with unit non-response under deterministic non-response model. Later, a concept of random non-response was studied by Singh and Singh (1978, 1985), Singh and Joarder (1998) among others. Alternatively, a situation when item non-response present in data set, a technique known as imputation was suggested by Rubin (1976) to mitigate the problem

of missing data.

In **Chapter 1**, we provide an introductory review on the theory of non-response, imputation and random non-response. Some relevant and important results are stated in this chapter. This chapter also consists of the basic definitions and notations relevant to this thesis.

In **Chapter 2**, we have proposed various Searls-type ratio and difference imputation methods on the lines of Ahmad et al. (2006). It is a well known fact that the optimal ratio type estimator attains the MSE of regression estimator (or optimal difference estimator) but while using Searls-type transformation (1964) this may not always happen. These imputation methods are shown to perform better than the imputation procedures of Ahmed et al. (2006). These difference imputation methods also perform better than the imputation procedure of Diana and Perri (2010). The proposed ratio imputation methods may even outperform the proposed difference imputation methods under certain optimality condition. This study is concluded with the numerical study alongside the theoretical comparison.

Chapter 3 introduces some improved methods of imputation using higher order moment of an auxiliary variable while imputing missing values. The performance of the proposed imputation methods are investigated relative to the estimators proposed by Mohamed et al. (2017) and by Bhushan and Pandey (2016). A comparative study has been carried out and it has been shown that the proposed estimators perform better in comparison to estimators proposed by Mohamed et al. (2017) and Bhushan and Pandey (2016). The theoretical findings are supported by an empirical study on real populations and a simulation study using hypothetical situation.

In **Chapter 4**, we propose some new imputation methods by extending the work of Bhushan and Pandey discussed in chapter 2 and chapter 3 using multi-auxiliary information. The popularly used imputation like mean imputation, ratio method of imputation, regression method of imputation and power transformation method are special cases of the proposed methods apart from being less efficient than the proposed methods. The proposed imputation methods can be considered as an efficient extension to the work of Singh and Deo (2003), Singh (2009), Ahmed et al. (2006), Diana and Perri (2010) and Bhushan and Pandey (2016). The theoretical results are derived and comparative study is conducted and

the results are found to be quite encouraging providing the improvement over the all the discussed works.

In **Chapter 5**, we have proposed some efficient estimators of population mean in presence of non-response. These estimators are suggested for both single phase sampling and two phase sampling. The estimators of mean obtained from proposed technique remains better than the estimators obtained by Cochran (1977), Khare and Srivastava (1995), Rao (1983, 1986) and Singh and Kumar (2008, 2010) under the derived optimality condition. The mean squared error of the resultant estimators is found to be less than that of the MSE of the estimators proposed Cochran (1977), Khare and Srivastava (1995), Rao (1983, 1986) and Singh and Kumar (2008, 2010) in an empirical study conducted on three populations.

In **Chapter 6**, we propose some new classes of estimators of population mean under non-response using bivariate auxiliary information. We have considered two auxiliary variables which are more efficient for the study character. It is observed that the proposed estimators are better perform as comparison to conventional estimators proposed by Singh and Kumar (2010b), Sabbir and Khan (2013) and Bhushan and Naqvi (2015) in order to support the results a comparative study is also carried out both theoretically as well as empirically.

In **Chapter 7**, we have constructed some improved difference and ratio type estimators of the population mean under two different situation of random non-response considered by Tracy and Osahan (1994) using Searls (1964) philosophy. The proposed difference and ratio type estimators remain better than the estimators obtained by Singh et al. (2007) in presence of random non-response. A comparative study has been performed and obtained some optimality conditions under which the conventional conclusion is reversed. It has been shown that the proposed estimators perform better in comparison to conventional estimators.

Chapter 8 proposes some new estimators of the population variance in presence of random non-response based on Searls (1964) philosophy. The proposed estimators remain better than the estimators obtained by Ahmed et al. (2005) in presence of random non-response using auxiliary variables. A comparative study has been performed and conditions for optimality have been obtained. It has been shown that the proposed estimators perform better in comparison to conventional estimators.

Bibliography

[1] Ahmed, M. S., Titi, O. A., Rawi, Z. A. and Dayyeh, W. A. (2005): Estimation of finite population variance in presence of random non-response using auxiliary variables. *Infr. Mang. Sci.*, 16(2), 73-82.

[2] Ahmed, M.S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006): Estimation of a population mean using different imputation methods. *Statistics in Transition*, 7(6): 1247-1264.

[3] Bhushan S. and Naqvi N.(2015): Generalized efficient classes of estimators in presence of non-response using two auxiliary variables. *Jour. Stat. Mang. Syst.*

[4] Cochran, W.G.(1977): Sampling techniques, 3rd edn. New York: *John Wiley and Sons.*

[5] Diana, G. and Perri, P.F. (2010): Improved estimators of the population mean for missing data.*Comm. Statist. Theory Methods*, 39, 3245-3251.

[6] Khare, B.B & Srivastava, S.(1995): Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. Proc. *Indian Nat. Sci. Acad.*, 65, 195-203.

[7] Hansen, M.H. and Hurwitz W.N. (1946): The problem of non-response in sample surveys. *J. Am. Stat. Asso.*, 41, 517- 529.

[8] Mohamed, C., Sedory, S. A., Singh, S. (2017): Imputation using higher order moments of an auxiliary variable, Commun. Stat. Simul. Comput., DOI:10.1080/03610918.2016.1208235.

[9] Rubin R.B. (1976): Inference and missing data. *Biometrika*, 63(3), 581-592.

[10] Rao, P.S.R.S.(1983): Randomization Approach in Incomplete Data in Sample Surveys, eds W.G. Madow, I. Olkin, and D.B. Rubin, *Academic Press*, 2, 33-44, New York.

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

[12] Searls, D. T. (1964): The Utilization of a Known Coefficient of Variation in the Estimation Procedure. *J. Am. Stat. Asso.*, 59, 1225-1226.

[13] Singh S, Singh R (1979): On random non- response in unequal sampling. *Sankhya*, 41, 27-37.

[14] Singh S, Singh R (1985): On random non- response in double sampling with difference estimator. *Commun. Stat. Simul. Comput.*, 14, 747-757.

[15] Tracy, D. S. and Osahan, S. S. (1994): Ranom non-response on study variable versus on study as well as auxiliary varibles, *Statistica*, 54, 163-168.

[16] Singh, S. and Joarder, A. H. (1998): Estimation of finite population variance using random non-response in survey sampling, *Metrika*, 98, 241-249.

[17] Singh, S. and Deo, B. (2003): Imputation by power transformation. *Statistical Papers*, 44, 555-579.

[18] Singh HP, Chandra P, Joarder AH, SIngh S (2007): Families of estimators of mean, ratio and product of a finite population using random non-response. *Test*, 16, 565-597.

[19] 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.

[20] Singh, S. (2009): A new method of imputation in survey sampling, Statistics. *A Journal of Theoretical and Applied Statistics* , 43(5), 499-511.

[21] Singh, H.P & Kumar, S.(2010): Estimation of mean in presence of non-response using two phase sampling scheme. *Stat. Papers*, 51, 559-582.

[22] Singh, H.P & Kumar, S.(2010b): Improved estimation of population mean under the two phase sampling with sub sampling the non-respondents. *J. Stat. Plan. Infer.*, 140, 2536-2550.

[23] Shabbir J., Khan N. S.(2013): On estimating the finite population mean using two auxiliary variables in two phase sampling in the presence of non-response. *Comm. Statist. Theory Methods*, 42, 4127-4145.