Estimation of Mean Using Multi Auxiliary Information in Presence of Non Response

Sunil Kumar\textsuperscript{1,\text昝}, Housila P. Singh\textsuperscript{b}

\textsuperscript{a}Department of Statistics, University of Jammu
\textsuperscript{b}School of Studies in Statistics, Vikram University

Abstract

For estimating the mean of a finite population, three classes of estimators using multi-auxiliary information with unknown means using two phase sampling in presence of non-response have been proposed with their properties. Asymptotically optimum estimator(AOE) in each class has been identified along with their mean squared error formulae. An empirical study is also given.

Keywords: Study variate, multi-auxiliary variates, mean squared errors, two-phase sampling, non-response.

1. Introduction

In survey sampling, it is well established that the use of auxiliary information results in substantial gain in efficiency over the estimators which do not use such information. Out of many ratio, product and regression methods of estimation are good examples in this context. When the correlation between the study variate \( y \) and the auxiliary variate \( x \) is positive (high) the ratio method of estimation is quite effective. On the other hand if this correlation is negative (high) the product method of estimation envisaged by Robson (1957) and rediscovered by Murthy (1964), can be employed. In large-scale sample surveys, we often collect data on more than one auxiliary character and some of these may be correlated with \( y \). Estimators using information of the known population mean of an auxiliary variable have generalized to the cases when such information is available for more than one auxiliary variables by several authors as Olkin (1958), Raj (1965), Rao and Mudholkar (1967), Singh (1967), Srivastava (1971) and Mohanty and Pattanaik (1982) and Agrawal and Panda (1993) \textit{etc}. But in many situations of practical importance it has been observed that the population means of auxiliary variables are not known. So we use two-phase sampling scheme for estimating the population means of auxiliary variables. Srivastava (1981) suggested a class of estimators for estimating the population mean in two-phase sampling assuming that responses for all variables are available for each unit selected in the sample. But in practice, the problem of non-response often arises in sample surveys. In such situations for single variable survey, the problem of estimating the population mean using sub-sampling scheme has been first considered by Hansen and Hurwitz (1946). Further improvement in the estimation procedure for population mean in presence of non-response using auxiliary variable was considered by Cochran (1977, p.374), Rao (1986, 1987), Sarndal \textit{et al.} (1992, p.583), Khare and Srivastava (1993, 1995, 1997), Okafor (1996), Tabasum and Khan (2004, 2006), Khare and Sinha (2004, 2007), Singh and Kumar (2008a, b; 2009a, b) and Singh \textit{et al.} (2010).

\textsuperscript{1}Corresponding author: Lecturer, Department of Statistics, University of Jammu, J & K, India, E-mail: sunilbhougal06@gmail.com
In this paper we have suggested three classes of estimators for estimating the population mean of the study variate using multi-auxiliary information with unknown population means using two-phase sampling in presence of non-response. The expressions for bias and mean squared errors of the suggested classes of estimators have been derived. The conditions for attaining minimum mean squared errors of the proposed classes have also been investigated. An empirical study is given in support of the present study.

2. Sampling Procedure and Notations

Consider a finite population \( U = (U_1, U_2, \ldots, U_N) \) of \( N \) units. Let \( y \) denote the study character whose population mean \( \bar{Y} \) is to be estimated using information on \( p \) auxiliary variates \( x_1, x_2, \ldots, x_p \). Let \( y_{ij}, x_{1j}, x_{2j}, \ldots, x_{pj} \) denote the values of the variates \( y, x_1, x_2, \ldots, x_p \) respectively, on the \( j^{th} \) unit \( U_j \) of the population \( U, j = 1, 2, \ldots, N \). When the population means \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p \) of the auxiliary variates \( x_1, x_2, \ldots, x_p \) respectively are known several multivariate ratio and product estimators of the population mean \( \bar{Y} \) have been formulated along with their properties for instance see Olkin (1958), Raj (1965), Rao and Mudholkar (1967), Singh (1967), Srivastava (1971), Mohanty and Pattanaik (1982) and Agarwal and Panda (1993). The population is supposed to be divided in \( N_1 \) responding and \( N_2 \) non-responding units such that \( N_1 + N_2 = N \). However, in certain practical situations population means \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p \) of auxiliary variates \( x_1, x_2, \ldots, x_p \) respectively are not known a priori in which case the technique of two-phase (or double) sampling can be useful. In two-phase sampling a first phase sample of size \( n' \) is drawn from the population by simple random sampling without replacement(SRSWOR) scheme on which only the auxiliary variates are measured in order to furnish the good estimates of \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p \). A smaller second phase sample of size \( n < n' \) is selected from \( n' \) by simple random sampling without replacement(SRSWOR) and the study variate \( y \) is measured on it. Let \( (\bar{x}_1', \bar{x}_2', \ldots, \bar{x}_p') \) be the sample mean of the auxiliary variates \( x_1, x_2, \ldots, x_p \) respectively based on first phase sample of size \( n' \). Further let \( \bar{y} \) and \( (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p) \) be the sample means of the study variate \( y \) and auxiliary variates \( x_1, x_2, \ldots, x_p \) obtained from the second phase sample of size \( n \) when there is no non-response (i.e. complete response) in the second phase sample. In such situations the formulation of two-phase (or double) sampling multivariate ratio and product estimators can be done easily just replacing \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p \) by \( (\bar{x}_1', \bar{x}_2', \ldots, \bar{x}_p') \), see Srivastava (1981). If, however, there is non-response in the second phase sample, take a sub-sample of the non-respondents and re-contact them.

We assume that at the first phase, all the \( n' \) units supply information on the auxiliary variates \( x_1, x_2, \ldots, x_p \). From the second phase sample of \( n \) units, let \( n_1 \) units supply information on the study variate \( y \) and \( n_2 \) units refuse to respond. From the \( n_2 \) non-respondents, using Hansen and Hurwitz (1946) procedure we again select a sub sample of size \( m = n_2/k, (k > 1) \) units using SRSWOR assuming all the \( m \) units respond. Here we have \( n_1 + m \) responding units on the study variate \( y \) and consequently the estimator for population mean \( \bar{Y} \) using sub sampling scheme envisaged by Hansen and Hurwitz (1946) is defined by

\[
\bar{y}^* = \left( \frac{n_1}{n} \right) \bar{y}_{(1)} + \left( \frac{n_2}{n} \right) \bar{y}_{(2)},
\]

(2.1)

where \( \bar{y}_{(1)} \) and \( \bar{y}_{(2)} \) denote the sample means of the study variate \( y \) based on \( n_1 \) and \( m \) units respectively. It is well known that the estimator \( \bar{y}^* \) is unbiased estimator of the population mean \( \bar{Y} \) and has the variance

\[
\text{Var}(\bar{y}^*) = \left( \frac{1 - f}{n} \right) S_0^2 + \frac{W_2(k - 1)}{n} S_{0(2)}^2,
\]

(2.2)