A GENERAL PROCEDURE FOR ESTIMATING THE MEAN OF A SENSITIVE VARIABLE USING AUXILIARY INFORMATION

Tanveer A. Tarray*[1] and Housila P. Singh**
*Department of Mathematics, Islamic University of Science and Technology, Awantipora, Pulwama, Kashmir - India,192122
**School of Studies in Statistics, Vikram University, Ujjain – India, 456010.

ABSTRACT
This paper proposes classes of estimators for the mean of sensitive variable utilizing information on a non – sensitive auxiliary variable. Expressions for the biases and mean square errors of the suggested estimators correct up to first order of approximation are derived. It has been shown that the suggested new class of estimators based on the known population mean $\overline{X}$ and variance S_x^2 of the auxiliary variable X are better than conventional unbiased estimators which do not utilize the auxiliary information, Sousa et al.'s (2010) ratio estimator and Gupta et al.'s (2012) regression estimator under a very realistic condition.

KEYWORDS: Class of estimators, Randomized response technique; Mean Square error, Bias, Auxiliary variable.

MSC: 62D05.

RESUMEN
Este trabajo propone clases de estimadores para la media de una variable sensitiva usando información sobre un variable auxiliar no – sensitiva. Expresiones para los sesgos y errores cuadráticos medios de los estimadores sugeridos, corregidos hasta el primer orden de aproximación, son derivados. Se demuestra que la nueva clase de estimadores sugerida, basada en al conocimiento de la media $\overline{X}$ y varianza S_x^2 de la variable auxiliar X, son mejores que estimadores insesgados tradicionales que no usan la información auxiliar, como el estimador de razón de Sousa et al. (2010) y el estimador de regresión de Gupta et al. (2012) bajo condiciones muy realísticas.

1.INTRODUCTION

In survey sampling, it is well recognized that the use of auxiliary information results in substantial gain in efficiency over the estimators which do not utilize such information. Ratio, product, regression estimators and their many rectifications have been discussed in the literature. In survey research, direct reliable observation on the variable of interest Y is sometimes not possible because the variable may be sensitive in nature such as habitual tax evasion, reckless driving, indiscriminate gambling, abortion, etc. However we may be able to directly observe a highly correlated auxiliary variable X, for instance see Sousa et al. (2010, 2012) and Koyuncu et al. (2014). Eichhorn and Hayre (1983), Bar – Lev et al. (2004), Perri (2008) and many others have estimated the mean of a sensitive variable when the study variable is sensitive and there is no auxiliary variable. Sousa et al. (2010), Gupta et al. (2012), Koyuncu et al. (2014) and Tarray and Singh (2014) have proposed mean estimators based on randomized response technique (RRT) models in the presence of an auxiliary variable that can be observed directly.

In this paper we have made an effort for developing the classes of estimators of the population mean $\overline{Y}$ of the sensitive variable Y using two different situations: (i) when the population mean $\overline{X}$ of the auxiliary variable X is known; and (ii) when both population mean $\overline{X}$ and variance S_x^2 of the auxiliary variable X are known.

Let Y be the variable under study, a sensitive variable which can't be observed directly. Let X be a non – sensitive auxiliary variable that have a positive correlation with the study variable Y. Let S be a scrambling variable independent of the study variable Y and the auxiliary variable X. The respondent is asked to report a scrambled response for Y given by $Z_a = Y+S$, but is asked to give a true response for the auxiliary variable X. To obtain the second response, Hussain (2012) advocated the use of

[1] [1]Corresponding Authors e-mail ID: tanveerstat@gmail.com