A SIMULATION PROCEDURE ON THE APPROPRIATE SEED SELECTION METHOD IN RESPONDENT DRIVEN SAMPLING

Hasani I Pathirana* and Asoka Ramanayake K. P.
Department of Statistics, University of Colombo, Sri Lanka
hasani1991
pathirana@gmail.com

Respondent Driven Sampling (RDS) which is in the fast growing phase with the experimenting platform is currently the best method to reach "hidden" or "hard to reach" populations. Application of RDS is predominantly employed when the sampling frame is not available. Seed selection method, which is with an utmost importance in the data collection procedure of RDS has become a popular topic in the relevant literature with several methods of selections. Among them, random seed selection and selection weighting against the network size have been used frequently. In this research these two methods are the main concerns. By considering an adjusted version of the publicly available Project 90 data of original Colorado Springs study as the population for the purpose of this study and proportion of females in the population as the parameter of interest, data were generated through a simulation study using R software and for calculation of estimates and other analysis, the RDS package in R was incorporated. Simulations were performed for particular seed selection by varying number of seeds from 4 to 14 in increments of 2 and fixing number of coupons and waves at 3 and 6 respectively. For each such combination, 1000 samples were generated and estimates were obtained with respect to the three mostly used RDS estimators, Salganik-Heckathorn (S-H), Volz-Heckathorn (V-H) and Successive Sampling (SS). Finally the results of this effort elaborate that probabilistic seed selection method outperforms the random seed selection method and SS estimator outperforms the other two estimators for this particular data set. On top of all other findings, this study has developed a procedure to resample RDS samples and a simulation code to compare these two methods using any population with any parameter of interest. It has also shed light on the importance of carrying out formative research in the area of RDS.

Keywords: *Respondent Driven Sampling, Seed selection, Simulation, Resample, RDS*