Oral presentation: 125

Optimal parameter values for *Respondent-Driven Sampling*

N. J. Kularathna[1]* and K. P. A. Ramanayake

Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka
*naveenjkularathna@gmail.com

Respondent-Driven Sampling (*RDS*) is a new link tracing sampling technique which is used to collect data from hard to reach or hidden populations such as sex workers, *HIV* infected people, homeless people, etc. In *RDS*, individuals recruit other individuals through their social networks. The major benefit of *RDS* among other link tracing methods is that it achieves a probabilistic sample with known selection probabilities. Deciding on the number of seeds, coupons and waves are crucial prior to implementing a *RDS* study. However, since there are no universally accepted values for these parameters in *RDS*, they need to be determined based on the study. This research focused on finding optimal number of seeds, coupons and waves that give the highest level of accuracy for *RDS* estimates when all the parameters are free to change simultaneously. A publicly available partial dataset from Project 90, Colorado Spring study was used as the population. The simulation study used the most frequently used sets of values for the number of seeds, coupons and waves based on literature. As a result, 125 combinations of seeds, coupons and waves were formed and for each such combination, 1000 resamples were drawn from the population. The successive sampling estimator was used in this study to estimate the population parameters as it has been shown that it substantively outperforms all other estimators in *RDS*. The simulation results revealed that the estimated values converge to the true parameter value as the number of seeds and the number of waves increase and when the number of coupons decrease (up to 2). Once the sample size and the number of seeds have been determined, the proposed simulation process can be used to find the optimal number of coupons which gives the highest accuracy for any considered population characteristic.

Keywords: *Respondent-Driven Sampling* (*RDS*), optimal coupons, optimal seeds, optimal waves