Model-based methods for missing data in surveys with post-stratification information

Sahar Z Zangeneh*

Fred Hutchinson Cancer Research Center, Seattle WA, U.S.A., saharzz@fhcrc.org

Roderick J.A. Little

University of Michigan, Ann Arbor, MI, U.S.A., rlittle@umich.edu

We study maximum likelihood estimation of the population mean for a survey experiencing unit nonresponse, i.e., when a sampled unit does not respond to the entire survey. We consider situations where post-stratification information is externally available for the population. Without external information, unit nonresponse, may lead to missing-data (MD) mechanisms that are *Missing Not at Random* (MNAR), which generally require a model for the missing-data mechanism. We show that when the response mechanism is governed by a post-stratifier, incorporating such information into the likelihood enables us to fit models to data that are MNAR to obtain maximum likelihood estimates without the need to model the MD mechanism. This framework is then extended to also incorporate covariate information that is fully observed for the sampled units. We compare and contrast the proposed model-based methods to existing design-based methods empirically for incomplete categorical data.