

Abstract:

Multiple imputation (MI) for handling the problem of missingness in categorical variables has received considerable attention specially when the data structure is multilevel. The formulation of latent normal variables for representing incomplete categorical variables in the imputation model is increasingly being practiced by the developers of multiple imputation software packages. The aim of this paper is to evaluate the performances of three different multilevel multiple imputation models that use latent normal variables for dealing with incomplete categorical variables. We study two joint multivariate normal (MVN) imputation models and one fully conditional specification (FCS) imputation model that can be implemented in available software packages. This paper discusses the following: a) the effect of the cluster size, the number of clusters, the overall sample size and the intraclass correlation on the missing data handling method; b) the effect of the method of treating fully observed covariates in the joint MVN imputation model (as responses or predictors); c) the effect of the latent normal variable formulation method on the performance of the MI; and d) whether to choose joint MVN or FCS to deal with multivariate missingness. We present the results by conducting a simulation study with 1000 artificial educational data sets. Multiple imputations are conducted in Mplus, R-jomo and Blimp software packages.