



Imputation of multivariate continuous data with nonignorable missingness

Thais Paiva*

Universidade Federal de Minas Gerais, Belo Horizonte, Brazil - thaisvip@gmail.com

Jerry Reiter

Duke University, Durham, USA - jerry@stat.duke.edu

Regular imputation methods have been used to deal with nonresponse in several types of survey data. However, in some of these studies, the assumption of missing at random is not valid since the probability of missing depends on the response variable. We propose an imputation method for multivariate data sets when there is nonignorable missingness. We fit a truncated Dirichlet process mixture of multivariate normals to the observed data under a Bayesian framework to provide flexibility. With the posterior samples from the mixture model, an analyst can alter the estimated distribution to obtain imputed data under different scenarios. To facilitate that, we developed an R application that allows the user to alter the values of the mixture parameters and visualize the imputation results automatically. We demonstrate this process of sensitivity analysis with an application to the Colombian Annual Manufacturing Survey.

Keywords: missing data; missing not at random; ignorability; mixture model.