

Multiple imputation with diagnostics: opening windows into the black box

Yu-Sung Su^{1,*}, Gelman Andrew¹, Jennifer Hill²

1. Columbia University

2. New York University

* Contact author: yusung@stat.columbia.edu

Keywords: missing data imputation, model checking

Our new R package, “mi,” has several features that allow the user to get inside the imputation process and evaluate the reasonableness of the resulting model and imputations. These features include: flexible choice of predictors, models, and transformations for chained imputation models; binned residual plots for checking the fit of the conditional distributions used for imputation; and plots for comparing the distributions of observed and imputed data in one and two dimensions. In addition, we use Bayesian models and weakly informative prior distributions to construct more stable estimates of imputation models. Our goal is to have a demonstration package that (a) avoids many of the practical problems that arise with existing multivariate imputation programs, and (b) demonstrates state-of-the-art diagnostics that can be applied more generally and can be incorporated into the software of others.