Christiana Drake
Department of Statistics, University of California, Davis
e-mail: cmdrake@ucdavis.edu
Travis Loux
Department of Statistics, University of California, Davis

Not Missing at Random and Combined Odds Ratios from Mixture Models

Longitudinal studies and surveys often deal with incomplete observations. The validity of inference depends on the missingness mechanism [Little J.A, and Rubin, D.B., 2002]. When the missing data mechanism depends on observed data only, estimation of means and/or regression coefficients requires adjustment but is possible without further information. If the missingness mechanism depends on unobserved data, unbiased estimation requires further information. The information from random sub-samples of subjects whose responses are obtained, can be used to model the data using selection, shared parameter or pattern mixture models [Allison, 1994], which are identifiable in this case. However, the parameters obtained may not be the ones of interest to an investigator. A separate regression fit to responders and nonresponders will result in two regression coefficients when a single coefficient for the whole population is of interest. Multiple imputation [Rubin, D.B. 1987, Glynn etal, 1993] can lead to standard statistical analysis. Very large surveys can have more than 50% non-response. A naive approach using multiple imputation results in data sets with more than 50% imputed values. We will discuss logistic regression for a mixture model and compare it to multiple imputation when missingness depends on the unobserved data., The methods are illustrated with the Project Talent data set. The original survey was very large and baseline information is available for all participants. Study attrition exceeds 50% but random sub-samples of nonrespondents have almost complete follow-up.

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis with Missing Data, 2nd edition. New York: John Wiley

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys, New York: John Wiley

Glynn, R., Laird, N., and Rubin, D.B. (1993), The Performance of Mixture Models for Nonignorable Nonresponse With Followups. Journal of the American Statistical Association, 88: 984-993.