Variable selection for multiply-imputed data with application to dioxin exposure study

Qixuan Chen

Department of Biostatistics
Columbia University

Tuesday, February 18, 2013
3:00–4:00pm, New PI 6th Floor Multipurpose Room (6602)
Light refreshments provided

Abstract

Multiple imputation (MI) is a commonly used technique for handling missing data in large-scale medical and public health studies. However, variable selection on multiply-imputed data remains an important and longstanding statistical problem. If a variable selection method is applied to each imputed dataset separately, it may select different variables for different imputed datasets, which makes it difficult to interpret the final model or draw scientific conclusions. In this paper, we propose a novel MI-lasso variable selection method as an extension of the lasso method to multiply-imputed data. The MI-lasso method treats the estimated regression coefficients of the same variable across all imputed datasets as a group and applies the group lasso penalty to yield a consistent variable selection across multiple-imputed datasets. We use a simulation study to demonstrate the advantage of the MI-lasso method compared with the alternatives. We also apply the MI-lasso method to the University of Michigan Dioxin Exposure Study to identify important circumstances and exposure factors that are associated with human serum dioxin concentration in Midland, Michigan. This is a joint work with Sijian Wang.

Biographical Note

Qixuan Chen received her PhD in Biostatistics from the University of Michigan in 2009 and is now an Assistant Professor in the Department of Biostatistics at Columbia University. Her research interests involve the development of statistical methods for complex survey data and data with missing values. She is also interested in novel application of statistical methods to biomedical research. Her applied interests are broad, including mental health, epidemiology, environmental health science, health policy and management, and medicine.

---

[1] The PI Biostatistics Seminar Series is held on Tuesdays at New York State Psychiatric Institute. If you are interested in receiving regular announcements for our seminars in the future, or if you need further information, please contact Jina James (jamesji@nyspi.columbia.edu, (646) 774-7929).