Three One-Day Short-Courses at the 31st New England Statistics Symposium

Friday, April 21, 2017, University of Connecticut
8:30am — 5pm, at Rome Ballroom
http://ness.stat.uconn.edu/short-courses

Course 1: Fitting Mixed-Effects Models Using the Julia Language

Instructor Dr. Douglas Bates is Emeritus Professor of Statistics at the University of Wisconsin–Madison. His research interests are in the theory and practice of mixed-effects modeling, including the development of software to fit such models. A member of the R Core Development Team and the JuliaStats organization, he has co-developed the `lme4` package for R and the `MixedModels` package for Julia. He is a Fellow of the American Statistical Association.

Outline The purpose of the course is two-fold: to introduce the Julia programming language (`http://julialang.org`) as used in statistical computing and to describe the formulation and fitting of mixed-effects models. Julia is a relatively young language, similar in structure to R and Matlab/Octave, but more flexible and capable of much greater performance. The blog posting, `http://www.oceanographerschoice.com/2016/03/the-julia-language-is-the-way-of-the-future/`, describes some of these advantages.

Mixed-effects models include linear mixed-effects models (LMMs), generalized linear mixed-effects models (GLMMs) and nonlinear mixed-effects models such as those used in population PK/PD modeling. Some specialized forms are called multilevel models or hierarchical linear models, item-response models, and panel data models. One of the major recent advances in mixed-effects modeling is the ability to fit models with crossed random effects such as effects for "subject" and for "item". Theoretical and computational advances include reformulation of the log-likelihood of such models in a compact, easily evaluated form through sparse and/or partitioned matrix formulations.

Prerequisites Introductory linear models and linear algebra plus some experience with an interactive computing environment such as R. The course will introduce Julia through analogy to R.

Course 2: Practical Integrative Statistical Learning: Recent Developments and Case Studies


Instructors Dr. Kun Chen is Assistant Professor in the Department of Statistics, University of Connecticut. Dr. Chen's research focuses on multivariate statistical learning, high-dimensional statistics, and health informatics with large-scale heterogeneous data. Recently his project on integrative multivariate analysis with multi-view data is funded by the National Science Foundation. He has extensive interdisciplinary research experience in a variety of fields including insurance, ecology, biology, agriculture, medical imaging, and public health.

Dr. Robert Aseltine is Professor and Chair in the Division of Behavioral Sciences and Community Health and Deputy Director of the Center for Public Health and Health Policy at UConn Health Center. Dr. Aseltine is a medical sociologist with diverse research interests that include health disparities, suicide prevention, and the development of innovative medical and public health information systems. Over the past 20 years he has led several major studies funded by the National Institute of Mental Health, the National Institute for Alcohol Abuse and Alcoholism, the Substance Abuse and Mental Health Services Administration, and the Department of Defense. He currently serves on the Advisory Board of the Connecticut All-Payer Claims Database and is Vice Chair of the New England Comparative Effectiveness Public Advisory Council (CEPAC).

Drs. Chen and Aseltine are closely collaborating on data-driven suicide prevention studies through integrating big data from disparate sources including health care providers, insurance companies, and government.

Outline This short course focuses on practical predictive modeling and statistical learning techniques for analyzing large-scale heterogeneous data. In many fields, measurements of several distinct yet interrelated sets of characteristics pertaining to a single set of subjects and possibly collected from an array of sources, has become increasingly common. For example, individual health data may come from insurance claims, pharmacy visits, clinical records, patient surveys, and government statistics. The availability of such complex data makes tackling many fundamental scientific problems possible through integrative statistical learning, which is undergoing exciting development and is pushing for a genuine refinement of the conventional multivariate learning toolkit. In this course, several classes of interpretable predictive modeling techniques for simultaneous dimension reduction, feature construction and model estimation will be introduced. Practical case studies in health informatics regarding suicide prevention, drug abuse, race and ethnic disparities in health outcomes, etc, together with examples from insurance, finance and industrial engineering will be discussed. The course consists of 4 modules: 1) overview of integrative statistical learning and health informatics; 2) dimension reduction techniques with case studies; 3) integrative predictive modeling techniques with case studies; 4) more recent developments on data fusion and demonstrations with R.

Prerequisites Entry level graduate courses in statistics or exposures to statistical modeling are desirable. Participants are encouraged to bring their own laptop computers to the session and to have the latest versions of R installed on their computers. The participants will have the opportunity to go through examples using a new R package developed by the instructor.

Course 3: Subgroup Analysis and Treatment Scoring with Application in Precision Medicine

Instructor Dr. Menggang Yu is Professor of Biostatistics and the Director of the Biostatistics Shared Resources of the Carbone Cancer Center at University of Wisconsin–Madison. Dr. Yu has extensive expertise in clinical biostatistics, risk prediction, causal inference, and treatment selection. He developed strong interests in comparative effectiveness research when working with the General Practice Research Database (GPRD) and The Health Improvement Network (THIN) databases to replicate clinical trial results. He is the Principal Investigator of a Patient Centered Outcome Research Institute (PCORI) methodology grant examining innovative methods to match medically complex patients to treatments and interventions

Outline In the case of substantial heterogeneity of treatment effectiveness, a key aspect of medical decision making is on matching patients with the most effective treatments to improve treatment efficacy and adherence, which in essence is personalizing treatments. Exploratory subgroup analysis allows the identification of clinically relevant subgroups from pre-specified variables for such purpose. In this short course, we will introduce a general framework that encompasses many recent statistical methods for identifying subgroups of patients who may benefit from different available treatments. Compared with the traditional outcome-modeling approaches, these methods focus on modeling interactions between the treatments and covariates while by-pass or minimize modeling the main effects of covariates because the subgroup identification only depends on the sign of the interaction. Under the proposed framework, we may also estimate the magnitude of the interaction, which leads to the construction of scoring system ranking the personalized treatment effect. The proposed methods are quite flexible and can be used for analysis of both randomized clinical trials and observational studies. They also allow incorporation of regularization in face of high-dimensional data. We will examine the empirical performance of several procedures belonging to the proposed framework through extensive numerical studies and real data analyses. Besides basic concepts and general issues in analysis, the course will cover active research in modeling multiple and longitudinal outcomes and in meta-analysis.

Prerequisites The course is accessible to anyone with a knowledge of statistical inference at the level of introductory graduate level courses in mathematical statistics and probability. Exposure to causal inference (based on the potential outcomes) and statistical learning theory (e.g., regularization method) can be helpful, but is not required.