2014 UT Summer Statistics Institute

Division of Statistics + Scientific Computation, The University of Texas at Austin

James Scott

Multivariate Data Analysis with R

(Course Description)

This is a highly practical, intermediate-level course that emphasizes learning by doing on real data sets.  The course is organized as a series of eight hands-on R vignettes.  Each one will be anchored by a single data set and research question (prototypical of their genre), and a statistical technique for answering the research question.  For each case study, I will describe the intuition behind the method, the assumptions and limitations of the method, the implementation of the method in R, and the interpretation of the output.  (We’ll all do these last two together.)  Because I will focus more on data analysis and less on mathematical niceties, I will also compile a reading list for each topic.  That way you can learn a bit more detail, or remind yourself about a model many months down the road (when it actually comes time to put these ideas into practice).

Day 1:

1.     Group-wise models: partitioning variation among categories

2.     Multiple regression for continuous outcomes

Day 2:

1.     Models for binary outcomes

2.     Models for ordered and unordered categorical outcomes

Day 3:

1.     Models for event-count data

2.     Survival/hazard models

Day 4:

1.     Hierarchical models I

2.     Hierarchical models II