Preliminaries and Logistics

MATH 456 - Spring 2016

1/25/16

Science and Statistics

1. Scientific questions
2. Study Design
3. Data Collection
4. Data Entry
5. Data Management, Screening and Transformation
6. Multivariate Analysis

Course Content Learning Outcomes

Upon successful completion of this course, students will be able to:

- Translate a research question into an appropriate statistical analysis plan.
- Prepare data for analysis by cleaning and transforming raw data.
- Perform research in a reproducible manner.
- Build a multivariate statistical model used to examine a real process.
- Report the results of the analysis in plain language to a consulting client or collaborator.

Topics Covered

- Data cleaning
- Multiple linear regression
- Model building: Variable selection
- Dimension Reduction: Principle components and Factor Analysis
- Logistic regression
- Discriminant Analysis
- CART / Random Forests
- Predictive model building
- Missing Data

Structure of this class

- Quasi flipped classroom.
- Students read the book, review lecture notes and work on practice coding prior to class.
- Class time is devoted to reviewing the material, answering questions, working examples, discussing interpretations and limitations of methods learned.
- Mondays will have either a quiz or homework submission.
- Expect to bring your laptop every day to class.

Class materials

- Class website: http://norcalbiostat.github.io/MATH456
- First stop for all class materials.
- Details on weekly topics can be found on the schedule.
- The syllabus covers course details such as grading, office hours and required materials.
- Shared Google Drive Folder: I will use this to store and share data, and possibly submit homework (TBD)
- Questions?

Helpful web resources for learning R

- R Programming blog http://rprogramming.net
- Quick-R http://www.statmethods.net/
- Cookbook for R http://www.cookbook-r.com/
- R Examples Repository http://www.uni-kiel.de/psychologie/rexrepos/index.html
- R Bootcamp Data visualization tutorial http://norcalbiostat.github.io/R-Bootcamp/labs/Data_Visualization_Tutorial_Full.html