ST 518: Data Analytics II

Course Overview

Data Analytics II

This course addresses methods for understanding and analyzing categorical and count data. These include:

- Contingency tables, proportions, risk, odds
- Logistic regression
- Log-linear regression
- Over dispersion in binomial and poisson variables
- Mixed effects in the generalized linear model setting

Data Analytics II

Other topics we will cover:

- Data imputation—methods for handling missing data.
- Prediction—using generalized linear models to predict new observations.
- Scaling up—methods for/issues with performing all of the above on large datasets.

Some Overarching Themes

In the context of all of the methods/techniques listed on the previous slides, we'll pay special attention to:

- Drawing inferences (making decisions based on quantitative information)
- Statistical vs. practical significance (how important/substantial are the results?)
- Checking assumptions/robustness (are the tools we use robust to departures from the assumptions underlying those tools?)
- Summaries of data analysis (communicating technical results and the caveats that may accompany them)

Some Examples

Here are just a few examples of the kinds of data problems you will learn about:

- Internet Marketing: which of 3 different website layouts leads to the most forward clicks to a subsequent webpage?
- Medicine: how is the presence/absence of tumors affected by increasing drug toxicity?
- Education: how is first year retention in college related to first year GPA, gender, race, ethnicity and other factors?
- Forestry: how is the density of salmon in streams related to forest thinning treatments in riparian (stream-side) areas?