Social Statistics II (Soc 723) Spring 2016
W 11:45-2:15 (Lab: Th 11:45-2:15), Soc-Psych 329

Professor: Scott M. Lynch
Office: 258 Soc-Psych. (Hours by appointment)
Email: scott.lynch@duke.edu
Teaching Assistant: John Bumpus (john.bumpus@duke.edu)

Objectives

This course is the second in the first-year graduate sequence in statistics. In the first course, you learned (or reviewed) common descriptive statistical methods as well as the basics of classical statistical inference for a variety of statistical methods. At the end of the course, you covered the basics of linear and generalized linear modeling.

This course provides in-depth coverage of the linear model, including the statistical theory underlying it, methods and software used to estimate it, methods to diagnose and correct problems with it, and methods to extend it. A key goal in linear modeling is to evaluate whether relationships between variables are causal, and so several weeks will be devoted to delineating the conditions under which the linear model can establish causality, as well as developing and comparing related, contemporary causal methods. Finally, the linear model forms the foundation for numerous additional methods, including regression methods for panel data, generalized linear models, analysis of time series data, and others. We will cover these methods as well (at varying levels of depth). By the end of the course, you should be able to conduct publishable quantitative analyses, as well as critically evaluate most contemporary published quantitative literature in sociology.

Format and Requirements

The course meets for three hours twice per week, once for lecture and once for lab. Class will follow a lecture format with discussion. The lab will involve applying methods covered in class to data using both Stata and R. Attendance and participation in both the lecture and the lab is expected.

Grades will be determined by a midterm exam (20%) and a series of homework assignments of varying point values (80%). Each item will be graded on a 5 point ordinal scale, roughly corresponding to a typical A-F grade: 4=Excellent (A+), 3=Very Good (A), 2=Good (B), 1=Fair (C), 0=Poor (D/F). Your final grade will be computed as a rounded weighted average of the midterm and the homework.

Readings

There are two required texts for the course. You should read the material before the day listed on the schedule and be ready to discuss it.

Course Schedule

Week 1 (Jan 20) K1-4, App. A-B

- Introduction
- Math review
- Probability theory review
- Maximum Likelihood estimation
- OLS model in matrix form and OLS and ML solutions

Week 2 (Jan 27) K5-6, 15

- Model Specification, Dummy Variables, Interactions, and Transformations
- Various (all) regression model calculations
- Model Assumptions

Week 3 (Feb 3) K12

- Multicollinearity
- Scale Construction (e.g., factor analysis)

Week 4 (Feb 10) K7-8

- Heteroscedasticity and WLS estimation
- Robust standard errors
- Discussion of models with built-in heteroscedasticity

Week 5 (Feb 17), K19, K21

- Correlated errors and GLS estimation
- Random effects modeling for clustered data
- ARiMA models
- Nonnormal errors and robust estimation
Week 6 (Feb 24)
  • Midterm Exam

Week 7 (Mar 2) K9-11, M1-4, M9
  • Endogeneity
  • Causal thinking
  • IV estimation

Week 8 (Mar 9) M5
  • Matching methods for causal estimation
  • Propensity score estimation for causal estimation

Week 9 (Mar 23) M6-8
  • Regression methods for causal estimation
  • Heckman selection

Week 10 (Mar 30) K18, M11
  • Fixed and random effects models for panel data

Week 11 (Apr 6) K16-17
  • Generalized linear models: dichotomous logit and probit
  • Ordinal logit and probit
  • Multinomial logit

Week 12 (Apr 13) prior readings, cont’d
  • Poisson regression modeling
  • Negative binomial regression
  • Zero-inflated models

Week 13 (Apr 20) Handout
  • Missing data