Political Science 271
Advanced Statistical Applications

Winter Quarter 2020
SSB 104, Tuesday 12PM-2:50PM

Molly Roberts
SSB 399
meroberts@ucsd.edu
Office hours by appointment: https://calendly.com/molly-e-roberts

Rachel Skillman
SSB 343
rskillma@ucsd.edu
Office hours: Monday 10AM-12PM

Prerequisites
Political Science 270 (or equivalent)

Overview
This course is the second course in the quantitative research methods sequence at the UCSD Political Science department. Building on Political Science 270, this course teaches advanced statistical tools for empirical political science. In the first half of the course, we will focus on techniques for model-based inference, with a specific focus on generalized linear models. We will cover the basics of the fundamental statistical principles underlying these models (e.g., maximum likelihood theory) as well as a variety of estimation techniques. In the second half of the course, we will focus on model selection, regularization, and missing data. The ultimate goal of this course is to provide students with adequate methodological skills for conducting cutting-edge empirical research in their own fields of substantive interest.
Assessment

Your grade will be based on a combination of:

- **Homeworks (40%)**: Four problem sets will be given throughout the quarter, skewed heavily toward the beginning of the quarter. Problem sets will contain analytical, computational, and data analysis questions. Each problem set will be counted equally toward the calculation of the final grade. The following instructions will apply to all problem sets unless otherwise noted.

  - Late submission will not be accepted unless you ask for special permission from the instructor in advance. Unless otherwise stated, all homework, including the proposal, is due before class time, that is by 12PM on Tuesdays. Problem set write-ups should be turned in electronically via a Dropbox link on the top of the problem set. All problem sets should be reproducible, in that they should include working code.
  
  - Working in groups is encouraged for conceptual and sometimes technical discussion, but each student must submit their own writeup of the solutions that shows their independent work on the assignment. In particular, you should not copy someone else’s answers or computer code. We also ask you to write down the names of the other students with whom you solved the problems together on the first sheet of your solutions. At times, the instructor will specify that for particular problems or problem sets that students should not work with others.
  
  - For analytical questions, you should include your intermediate steps, as well as comments on those steps when appropriate. For data analysis questions, include annotated code as part of your answers. All results should be presented so that they can be easily understood. We strongly encourage you to use R Markdown to document both your answers and code.

- **Midterm (25%)**: Students will complete an in-class midterm halfway through the quarter. This midterm will be set up to mimic questions that are frequently asked on comprehensive exams, in order to prepare students for the methods exam.

- **Final project (25%)**: The final project will be a poster which typically applies a method learned in this course to an empirical problem of your substantive interest.

  - I encourage you to work with another student on your poster. By co-authoring you will (1) learn how to effectively collaborate with someone else on your research, which is very important in political science where most cutting-edge research is collaborative and (2) more likely have a good, potentially publishable paper (multiple brains are usually better than one).

  - Unless you already have a concrete research project suitable for this course (e.g., from your dissertation project), we recommend that you start with replicating the results in a published article and then improve the original analysis using the methods learned in this course (or elsewhere). Oftentimes, the most time-consuming part of a research project is data collection (which is not the focus of this course) and using data someone has already archived for their publication and made publicly available gets around this problem.
• Students are expected to adhere to the following deadlines:

  – January 21: Turn in a brief description of your proposed project. By this date you
    need to have found your coauthor, acquired the data you plan to use, and completed a
    descriptive analysis of the data (e.g. simple summary statistics, crosstabs and plots).
  – March 17: Poster session: Class time will be spent in a poster session on this day where
    students present the results of their paper and comment on one another’s work.

• Participation and presentation (10%): Students are strongly encouraged to ask questions
  and actively participate in discussions during lectures and recitation sessions.

Academic Honesty and Plagiarism

All of your graded work must be done by you. If you are unfamiliar with the University’s policy on
academic integrity, please see http://students.ucsd.edu/academics/academic-integrity/policy.html.

Syllabus and Plan

The syllabus will be updated periodically throughout the course, so that we can keep with the
cadence of the class. I will post to Piazza when such updates are made.

Reading and Textbooks

We will read chapters from these books throughout the course. We recommend that you purchase
the King book. The others we will only read a few chapters from, and will be available on electronic
reserve.

• James, Gareth, Daniela Witten, Trever Hastie and Robert Tibshirani. *An Introduction to
  Statistical Learning with Applications to R*. Springer-Verlag, 2013. (available online at: http://www-bcf.usc.edu/~gareth/ISL/ISLR\%20First\%20Printing.pdf.)

• King, Gary. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.*


Piazza

We will be using Piazza for general discussion and questions and answers throughout the class.
Piazza allows students to see other students’ questions and learn from them as well as answer
them. Your respectful and thoughtful participation in the discussion forum will count toward
your participation grade. Please do not e-mail the instructor or TA with questions (post them on
Piazza!) unless they are personal in nature. We will check the Piazza forum daily to provide our
own answers and contributions.
Software

We will be using R, an open-source statistical package. You can download it from the web here: http://cran.r-project.org/

We highly recommend you submit your problem sets in RMarkdown.

COURSE SCHEDULE

1  January 7:  Course Introduction and Linear Regression Reframed

Reading
Chapter 1-3 King

2  January 14:  Basic Probability and Intro to Maximum Likelihood

Reading
Chapter 4.1-4.3, King
Chapter 1, Ward and Ahlquist

3  January 21:  Section During Class Time

Project proposal due

4  January 28:  Optimization, Uncertainty, and Properties of MLE

Problem Set 1 due

Reading
Chapter 4.3-4.6, King
Chapter 2 & 4, Ward and Ahlquist

5  February 4:  Binary Dependent Variables

Problem Set 2 due
Reading
   Chapter 5.1-5.2, King
   Chapter 3, Ward and Ahlquist

6 February 11: Simulation
Problem Set 3 due

Reading
   Chapter 5.2, King
   Chapter 6, Ward and Ahlquist

7 February 18: Count and Ordered Models
Problem Set 4 due

Reading
   Chapter 5.4-5.10, King
   Chapter 8, and 10, Ward and Ahlquist

8 February 25: Midterm

9 March 3: Prediction vs. Causal Inference & Model/Variable Selection

Reading
   Sections 2.1.1 and 2.1.2, Chapter 5, James et al.

10 March 10: Regularization and Missing Data

Reading
   Chapter 6, James et al.

11 March 17: Poster Session (11:30-2:30)