PS 231-A: Quantitative Analysis in Political Research

Prof. Sean Gailmard
Dept. of Political Science
734 Barrows
Tel: 510-642-4677
Email: gailmard@berkeley.edu

Fall 2014
UC Berkeley
Location: 202 Barrows
Time: M 9:00 am – 12:00 pm
Office hours: W 10:00 am – 12:00 pm

This is a first course on statistical inference and modeling for use in social science research. It covers the theory of statistical inference, essential concepts in statistical modeling, justifications for and problems with common statistical procedures, and how to apply procedures to empirical social science data to draw conclusions relevant to positive social theory. We will pay particular attention to the motivation for statistical inference and modeling from the standpoint of social science. Lectures and reading will primarily cover theory and simple examples. Problem sets will cover both simple theoretical extensions and applications of tools we develop to real data.

Required Skills. Students should have completed math camp or its equivalent, and passed the math camp exam. Students should have a working knowledge of arithmetic, algebra, and elementary calculus. The course is suitable for students with a large range of prior exposure to statistics and mathematics. Students with Ph.D.-level training in mathematical statistics from a statistics department will not find that it pushes their capabilities; students with less background than this should find at least some challenges, conceptual or technical. *All students capable of gaining admission to a Berkeley Ph.D. program can fully succeed in this class regardless of prior technical preparation other than the required skills listed above.*

Reading and Class Sessions

Readings come from two sources. One is required; for the other you have some discretion. The required reading is a textbook that I happen to have written, *Statistical Modeling and Inference for Social Science*. It might be available in some bookstore or other, but everyone uses Amazon anyway. For the other readings, students should pick based on their background with statistics and intention to specialize in quantitative methods. Students with a relatively basic background or desiring less technical coverage of concepts should choose readings from *Statistics*, Freedman, Pisani, and Purves (3rd or 4th ed. are both fine). While this is an introductory undergraduate text, it is relatively uncompromising conceptually and provides excellent explanations. Students desiring strong technical foundations should choose readings from *Statistical Inference*, Casella and Berger (1st or 2nd ed. are both fine). This is a very solid contemporary text on mathematical statistics. Of course, there is nothing stopping you from completing readings from all three sources, and students interested in more advanced coverage would probably find Freedman, Pisani, and Purves (*i*) relatively straightforward, and (*ii*) very enlightening conceptually.

Whether you complete readings before the associated lecture session or after is up to you. You should complete them in relatively close to the class session associated with a given topic in the readings.

Most of the class sessions will consist of lectures, and mostly on theory, concepts, and simple examples. In general, readings will cover more material than the associated lecture. Despite the lecture format the nature and goals of this class require active involvement, discussion, and participation from students.

Grades

The course grade will be determined as follows:

1. Problem sets (3-4 total): 60%
2. Participation: 15%
3. Final exam: 25%

1. There will be required problem sets roughly every two-three weeks throughout the term, each equally weighted in the final grade. Problem sets will help students develop facility with important theoretical concepts, as well as applications of techniques to datasets.

2. Class sessions, while primarily composed of lectures, will require extensive student involvement and participation. I will use the whole 15% range for these scores. If you are attentive and ask informed questions you will score close to the maximum. If I do not know what your voice sounds like you will score close to the minimum.

3. There will be a comprehensive take-home final exam at the end of the semester.

GSI and Discussion Section

The GSI is Katherine Michel, a Ph.D. candidate in the Department. Katherine will grade problem sets and exams, hold office hours to provide advice on concepts and problem sets, and will conduct a weekly discussion section. Her email address is katherine_michel@berkeley.edu.

The weekly discussion will consist of two parts: a computer software and data review, on how to execute in **R** (and possibly Stata) procedures covered in the week's lecture; and a conceptual review/translation, elaborating the theoretical material in the week's lecture. How long each part lasts in a given week depends on student interest and capability. Students are strongly urged to attend, and take a laptop to, the part of the section on software & data review, because this is the primary delivery of an important part of this class on how to actually execute statistical procedures with data.

Sequence of Topics and Readings

This is a sequence of topics and attendant readings but not a schedule. We will cover topics in this order but I have stopped trying to predict how long each topic will take. We will work on each one until we seem to understand it. NOTE. G refers to the required course text; FPP refers to Freedman, Pisani, and Purves (4th ed.); CB refers to Casella and Berger (2nd ed.).

Session 1 Descriptive Statistics.
- G, chapters 1-2
- FPP, chapters 3-4, 8-12

Session 2 Probability.
- G, chapters 3-4
- FPP, chapters 6, 13-14
- CB, chapter 1 (section 1.1-1.3)

Session 3 Random Variables and Probability Distributions.
- G, chapter 4
- FPP, chapter 15
- CB, chapters 1 (sections 1.4 - 1.6), 4

Session 4 Expectation, Variance, Covariance.
- G, chapter 5
- CB, chapter 2

Session 5 Families of Distributions.
- G, chapter 6
- CB, chapter 3

Session 6 Models.
- G, chapter 6
- CB, chapter 3

Session 7 Sampling, Sampling Distributions, and Central Limit Theorem.
- G, chapter 7
- FPP, chapters 16-19
- CB, chapter 5 (sections 5.1 - 5.2)

Session 8 Sampling; Derived Distributions.
- G, chapter 7
- FPP, chapter 20
- CB, chapter 5 (sections 5.3 - end)

Session 9 Hypothesis testing.
- G, chapter 8
- FPP, chapters 26-29
- CB, chapter 8

Session 10 Hypothesis Testing II.
- G, chapter 8

Session 11 Interval Estimation, Likelihood.

- G, chapter 9
- FPP, chapters 21 and 23
- CB, chapters 7, 9, 10

Session 12 Likelihood, point estimation.

- G, chapter 9
- FPP, chapters 1-2

Session 13 Estimation; evaluating estimators.

- G, chapter 9
- FPP, chapters 1-2