GOVT 6019
Introduction to Probability and Statistics

Tom Pepinsky
pepinsky@cornell.edu

Fall 2020

Class Meeting Times		Office Hours
Monday/Wednesday, 9:55-11:10		Mondays, 1-3pm (sign up)
Via Zoom and Baker 119		Via Zoom
Section Meeting Times	Teaching Assistant	TA Office Hours
Friday, 3-3:50	Rebaka Johnson	Tuesdays, 1-3pm (sign up)
Via Zoom and Room TBD	rj388@cornell.edu	Via Zoom

Course Description

This course is an introduction to core concepts in probability and statistics, with a focus on applications in political science. Assuming no background knowledge in either statistical reasoning or the fundamentals of probability, we will cover the building blocks for statistical inference. Topics will include variables and distribution, descriptive statistics, sampling, hypothesis testing, Bayes' rule, ordinary least squares regression, missing data, causal inference, and data visualization.

As part of this course, students will also learn basic scripting skills in the statistical programs Stata and R. We will use these two statistical packages for all assignments, building intuitions using a simulation-based approach. Students will learn to work with variables, scalars, vectors, and matrices; will gain experience in writing simple loops and functions; and will build skills by doing basic mathematical operations "the long way" and then checking their results using built-in functionality in each software package. We will emphasize full replicability in all assignments.

Course Objectives

By the end of this course, students will

1. understand basic terms and concepts in probability

2. have a good intuitive understanding of statistical reasoning

3. be familiar with the basics of inference (descriptive and causal)

4. be able to use R and Stata to "do statistics"

5. see how apply all of these skills to questions of politics

Required Materials

All course materials are available via Canvas.

Prerequisites

This course assumes only basic knowledge of high school-level algebra. No other mathematics, statistics, or computer science background is required.

Course Information

Class Structure

Owing to the challenges of COVID-19, this course wll be delivered in a "flipped classroom" model featuring asynchronous lectures with synchronous sections and computer labs, offered in a "hybrid" model that includes in-person option. Students may also participate remotely via Zoom.

Lectures

Pre-recorded lectures will be provided via Canvas. Each week, there will be 2 or 3 short videos that comprise the lecture, and you should plan to watch each prior to each Monday's regular section meeting. You may watch these at your leisure, as many times as you like. I will also make available PDF copies of the lecture slides, which you are free to download and save as a reference.

Regular Sections

Each Monday's scheduled course meeting will be a discussion/review section. I will provide a short overview of what you learned from the pre-recorded lectures that you watch, and will then open it up to the class for questions, comments, and discussion. If we do not need the full 75 minutes, we can break early.

Computer Labs

Each Wednesday's scheduled course meeting will be a computer lab session in which we will learn the basics of working with Stata and R. The .R and .do files that we will use in these lab sessions will also be available via Canvas.

Discussion Sections

Each Friday afternoon, your TA will lead a discussion section. This is your opportunity to ask more questions about the material that we have covered, and to check in with any questions about your homework assignments.

Assignments

Every Wednesday, you will get an assignment that builds your ability to use the statistical software (taught that day) to master lecture material (taught the previous Monday). It is due the following Monday. There are no other tests or exams.

The final assignment is create a statistical computing "cheatsheet:" a single document that contains all of the statistical commands that you learned in both softwares.

Course Flow: Example

- *Monday: Regular section.* Homework #2 due, Week #3 lectures must be finished

- *Wednesday: Lab meeting.* Homework #3 assigned, and Week #4 lectures made available

- *Friday: Discussion Section.* Review Week #3 materials.

Course Policies

I will not grant incompletes in this class except as a consequence of extraordinary personal circumstances. But these are COVID times, so if you find yourself facing extraordinary personal circumstances please do not hesitate to reach out to your TA, or to me.

Academic Integrity and Honesty

I expect all students to abide by Cornell University's Code of Academic Integrity. You must turn in work that is your own. However, I expect students to work together to solve problems and to reason through the assignments, and I strongly encourage collaboration. All assignments, moreover, are "open-book open-note open-browser."

Accommodations for Disabilities

In compliance with the Cornell University policy and equal access laws, I am available to discuss appropriate academic accommodations that may be required for student with disabilities. Students are encouraged to register with Student Disability Services to verify their eligibility for appropriate accommodations. Students seeking accommodations should submit to me an accommodation letter from Student Disability Services within the first two weeks of the semester.

Discrimination

Students and faculty each have responsibility for maintaining an appropriate learning environment. Students will treat one another with respect and courtesy. Discrimination based on race, gender, sexual orientation, religion, national origin, ability status, political affiliation, or any other personal characteristic will not be tolerated.

Religious Observance

I am happy to provide reasonable and timely accommodations for sincerely held religious beliefs. Please review the syllabus closely to determine if your religion will present any scheduling conflicts with any of the assignments. You must inform me of any conflicts within the first two weeks of the semester.

Course Schedule

Week 0, 09/02: **Introduction and Housekeeping**

Week 1, 09/07 and 09/09: **Variables**

- types of variables
- central tendencies

Week 2, 09/14 and 09/16: **Distributions**

- density and mass functions
- normal, binomial, uniform, various exotic ones

Week 3, 09/21 and 09/23: **Expectations and Variances**

- definitions and properties
- estimates

Week 4, 09/28 and 09/30: **Sampling**

- randomization
- the Weak Law of Large Numbers
- the Central Limit Theorem
- bias, consistency, and efficiency

Week 5, 10/05 and 10/07:   Confidence Intervals

- standard errors
- Z/t statistics
- the bootstrap

Week 6, 10/12 and 10/14:   Hypothesis Testing (*note: no lab on Wednesday*)

- null and alternative hypotheses
- t-tests
- tests of proportions
- one- and two-tailed tests

Week 7, 10/19 and 10/21:   Power

- false positives and false negatives
- power calculations
- effect sizes

Week 8, 10/26 and 10/28:   Bayes' Rule

- marginal, joint, and conditional probabilities
- sensitivity and specificity
- prior and posterior beliefs

Week 9, 11/02 and 11/04:   Functions of Multiple Variables

- correlation versus dependence
- contingency tables and χ^2 tests
- Pearson's and Spearman's ρ

Week 10, 11/09 and 11/11:   Bivariate Regression

- slope-intercept interpretation
- Gauss-Markov Theorem

Study Days, 11/16 and 11/18:   No class or lab

Thanksgiving, 11/23 and 11/25:   No class or lab

Week 11, 11/30 and 12/02:   Multiple Regression

- matrix representation

- "omitted variable bias"

- design and assignment

Week 12, 12/07 and 12/09:   Missing Data

- missing data mechanisms

- interval estimation and sharp bounds

- imputation

Week 13, 12/14 and 12/16:   Causal Inference

- the Fundamental Problem of Causal Inference

- potential outcomes

- directed acyclic graphs

- ignorability and unconfoundedness