ECON 3130: Introduction to Probability and Statistics

Cornell University, Fall 2022

Lectures: Mon/Wed, 8:05am–9:20am

Goldwyn Smith Hall G76-Lewis

Instructor

Douglas McKee (douglas.mckee@cornell.edu)
Office Hours: See Canvas

Teaching Assistants

Kalie Hwei-Gen Pierce (khp44@cornell.edu)
Section 201: Friday, 9:05am-9:55am, Rockefeller 115
Section 202: Friday, 10:10am-11:00am, Rockefeller 115
Office Hours: See Canvas

Leonardo Peñaloza-Pacheco (ljp96@cornell.edu)
Section 203: Friday, 1:30pm-2:20pm, Rockefeller 115
Section 204: Friday, 2:40pm–3:30pm, Rockefeller 115
Office Hours: See Canvas

Course Description

The Cornell Economics Department offers two alternative two-semester sequences in undergraduate econometrics. This course (Econ 3130) is the first in the more math-intensive sequence, and is an introduction to probability theory and statistics. These tools are most useful in the analysis of data to answer important substantive economic questions, but at the same time, probability theory is the way in which uncertainty is modeled across many fields of economics. For example, game theory relies on probability theory to describe the uncertain actions of agents in a game. Stochastic (i.e., random) factors are a key component of modern macroeconomic models. Health economists rely on probability theory to understand insurance markets, moral hazard, and adverse selection. Uncertainty about risk and return is a fact of life for economists studying financial markets.

In most econometrics classes, mathematical methods are introduced and then applied to a few examples. This class turns that around. We will focus on substantive questions first and then introduce mathematical methods that will help us answer them. The analysis will be conducted in a relatively rigorous manner, and we will formally prove several of the important theorems. I will also spend significant time giving you intuition behind these theoretical results. Learning how to apply these methods to the analysis of real data is also a key part of the course. This makes the utility of the methods obvious as well as giving you another way to understand the way these methods work.

The first part of this course will define the basic concepts of probability and the most common probability distributions. We will define properties of these distributions (e.g., mean, variance, cumulative distribution functions and moment generating functions) and develop methods for

inferring them. In the second part of the course we will then consider statistical methods that permit us to use small amounts of information (aka samples) to answer larger questions. Specifically, we will start with random sampling from a population, and derive the distributions of some sample statistics. We will also develop methods of estimating the parameters of random variables using data and quantifying the uncertainty inherent in our estimates. Finally, we will discuss hypothesis testing: the use of data to confirm or reject hypotheses we have about random variables in the population.

The prerequisites for this course are one introductory economics course (ECON 1110 or 1120) and familiarity with single variable calculus (MATH 1110 and 1120). We won't use what you learn in macroeconomics very much, but I will assume you have a working microeconomics vocabulary. And we will lean heavily on your knowledge of differentiation and integration (Math 1110) and you will see plenty of proofs that use sequences, Tayor series, and limits (Math 1120)

Class Time

Class time will be highly interactive, and each class will contain multiple group work activities. While you work, the teaching assistants and I will circulate around the room providing feedback and guidance.

1. **Application:** After a short lecture on a topic, you will work through problems similar to those you might find on an exam.
2. **Case Study:** You will discuss how a theoretical model might be applied to a real-world situation. This might be a case where we are trying to estimate a parameter (e.g., the variance of wages) using data or a case where we are modeling an uncertain quantity (e.g., the number of shoppers that enter a store on a given day).
3. **Invention:** You will attempt to solve a problem *before* a lecture on the topic. Studies have shown that students who do invention activities before learning a new method understand the method much more deeply than students that simply get a lecture on the method. They retain the knowledge longer and are able to apply the concepts more broadly. And with the right attitude, invention activities are a lot of fun.
4. **Sense-making:** You will move beyond black box thinking and interpret the meaning of the component parts of an expression or equation. This will often involve graphic visualization, and it is critical to deep understanding of the probablistic models and statistical estimators you'll learn in the class.

Please keep in mind that all of these activities are difficult and **struggle is expected!** You'll learn a lot more working through hard problems than easy ones.

During most classes, you will use the Poll Everywhere system to answer questions and give me feedback on what you are learning and what you are not. This keeps you engaged, and lets us know when we need to provide more direction and when we are ready to move on. You can answer questions in any web browser or use the Poll Everywhere app on your phone. Because Cornell has a university-wide license to this software, it's completely free to you.

Grades

Your grade for the class will be composed of four parts:

1. Problem Sets (20%)
There will be 6 problem sets during the semester. You must submit the problem sets as pdf's through the course web site by the due date. Your grade on each problem set will be based primarily on your answers to two randomly selected questions, but I will also expect you to submit answers to all the questions. Complete solution sets will be posted in the evening after the problem sets are due. It is your responsibility to read these solutions and make sure you understand them. Your lowest grade on a problem set will be dropped. **Late problem sets will not be accepted. Computer exercises must include the Stata .do and .log files that you produce.**

2. Quizzes (5%) At the end of most modules, you will take an online quiz on that module's material. Your grade will be your average score during the semester. Your lowest quiz will be dropped, and there will be no excuses for missing quizzes.
3. Class Participation (5%)
As you know from above, I believe there are major returns to attending class and participating in the activities. I also think it's easier to stay focused when the lecture is live. To further incentivize attendance and participation, 5% of your grade is based on a participation score equal to the fraction of classes where you clicked in for at least half of the Poll Everywhere opportunities. You can miss up to 5 classes and still get credit for participating in those classes, but again, this policy means there will be no excuses for missing additional classes.

If you cannot come to class for whatever reason, there is an additional way to earn the same participation credit: Within 24 hours of the end of class, you can watch the recorded lecture and submit answers to the questions I pose to students during the class. Along with your (mostly multiple choice) answers, you'll have to include explanations for your answers. You will get a lot more out of this exercise if you pause the recorded lecture as the questions come up and try to answer them right then.

4. First Prelim (20%)
Date: Thursday, September 29

5. Second Prelim (20%)
Date: Tuesday, October 25

6. Final Exam (30%)
Date: TBD

The final test will be given during finals period. The schedule will be posted at
https://registrar.cornell.edu/Sched/exams.html some time in early in the semester.

Exams are closed book, but you may bring one double-sided page of notes to the first prelim, two pages to the second prelim, and three pages to the final exam. You may use calculators during the exams.

We will also be doing *two stage exams* for the prelims in this class. You will first take the exam individually and hand in your test. Then you will take the exam in small groups, where each group works together and passes in one exam with their consensus answers. If your individual score is higher than your group's score, your grade on the exam will be your individual score. If your group's score is higher, your grade on the exam will be a weighted average of your individual score (90%) and your group's score (10%).

The main reason we conduct a second stage of the exam is to allow you to learn more during the exam. Traditional exams tend to be summative rather than formative, and two stage exams represent an opportunity to redress this imbalance. The process of discussing your answers with your teammates is a significant learning opportunity and supports the kind of collaborative learning that we encourage.

Final grades for the class will be determined by computing a weighted score based on the weights listed above. The weighted scores are assigned letter grades based on the following cut-offs:

Range	Letter
94–100	A
90–93	A-
87–89	B+
83–86	B
77–83	B-
71–76	C
65–70	C-
50–64	D
0–49	F

I expect this breakdown to result in about 45% of the class getting A-'s or A's, and reserve the right to make the cutoff's more generous if the exams are unexpectedly difficult. Rest assured that the cutoffs will not be made less generous under any circumstances. I will give A+'s to students who earn A's and show extraordinary mastery of the material by the end of the semester.

Excuses

Because one problem set is dropped, I do not consider excuses for missed problem sets. The only exception is prolonged/severe illness, which must be handled through the advising deans as per case (1) below.

With respect to exams, the Faculty Handbook lists four types of situations in which faculty are encouraged to make accomodations for missed work. However, the determination as to whether a particulary case warrants accomodation is ultimately the decision of the faculty member. Here is how the four cases are handled in this course:

1. Illness, or family or personal emergency: Any situations that fall under this category must be first brought up with the advising dean in the student's college. The advising dean

will then contact me directly, and I will make a determination based on the particular case. Do not email me directly about these issues.

2. Employment interviews. The student must provide me evidence of the interview and establish that (s)he has no control over the timing of the interview.

3. Religious observances. While I do my best not to schedule exams during religious holidays, please contact me at least two weeks in advance if an exam date/time conflicts with a religious holiday.

4. Athletics and Extracurricular Activities. Students in varsity athletics or recognized extracurricular activities must provide the standard permission slip from the staff responsible for the activity at least two weeks before the exam.

Final Exam Conflicts

There are two situations that I will consider for exam conflicts. First is a direct conflict where ECON 3130 and another class appear on the registrar's exam schedule at the same time. Second is 3 or more exams having a start time within 24 hours, as indicated on the registrar's exam schedule. If Exam 1 is on Monday at 2pm, Exam 2 is on Monday at 7pm, and Exam 3 is on Tuesday at 2pm, this is not more than 2 exams in 24 hours. If you have a conflict, you need to email me at least 2 weeks before the final exam, listing out the other classes involved and scheduled exam times. The date and time of the makeup is determined by me.

Grading FAQ

- **Are the tests cumulative?** The tests **are** cumulative. About 15% of the second prelim is on earlier material and up to half the final exam is on material covered on the first two prelims. In addition, you will need to use concepts from the earlier parts of the course in order to understand the later topics.
- **Is there extra work I can do to improve my grade?** No.
- **I didn't do as well as I had hoped early in the course. In determining my final grade, can you put more weight on the latter part of the course?** No.
- **I have X exams/assignments due within Y of each other. Can I reschedule the exam/hand in the homework late?** No. The first two prelim dates are given above, and the final exam will be posted in September. This gives you plenty of time to plan ahead. The only exception is more than 2 finals in 24 hours, see above.

Exam Regrades

While we take care to grade exams as fairly and consistently as possible, on rare occasions there may be grading mistakes. If you feel that your test has been graded incorrectly, you must submit it to the professor (not the TA), along with an explanation of the issue in writing. You must do this within 2 weeks of the exam being returned (not the date you pick it up) for it to be regraded. The entire exam will be regraded, and as a result it is possible for your grade to go down as well as up.

Academic Integrity

All students in the course are expected to follow the Cornell University Code of Academic Integrity and Cornell University Campus Code of Conduct. You can find the full text of these policies at the following links:

- Cornell University Code of Academic Integrity
- Cornell University Code of Campus Conduct

You are encouraged to study together and to discuss concepts covered in lecture with other students. You are also encouraged to discuss problem sets with your fellow students inside or outside your assigned small groups. You are welcome to meet in person or on Zoom.
Any work submitted by a student in this course for academic credit must be the student's own work. You can give "consulting" help to or receive "consulting" help from other students. However, this permissible cooperation should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an email, an electronic file, or a hard copy. Students are permitted to share their notes with other students who are also registered in the course this semester.

Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

When it comes time to take exams, you will need to complete these entirely on your own.
Please remember that academic integrity also applies to the proper use of course materials:

- Canvas is a restricted-access site. All materials posted to Canvas are prepared solely for students registered in this course during this semester.
- Students are not authorized to buy or sell or generally distribute course materials (i.e., course materials should not be shared with anyone not currently registered in the class). In particular, students are prohibited from buying and selling course materials through internet sites such as Chegg, CourseHero, and Slader.
- If you buy or otherwise access course materials through such a vendor, you face a charge of "Unauthorized Assistance," thereby violating the Code of Academic Integrity. Please note that Cornell faculty are able to trace posts from internet sites, including identifying the individuals who provide the original posts and those who read or download the posts.
- If you sell course materials without my authorization—including "derivative" materials such as your own class notes or other materials you prepare that are based on course materials—you are subject to a charge of "Academic Misconduct." You may also be participating in copyright infringement, as course materials are intellectual property that belong to the author and are not a student's property to sell.
- Please note that this prohibition applies to the videos that contain lecture material and section material—these should not be bought or sold or shared in any way with someone not enrolled in the class.

Note: The use of any materials prepared in a previous semester for this course, other than materials redistributed this semester, is strictly prohibited. As above, this includes "derivative" materials that former students might have prepared based on their course materials.

Text and Readings

The required textbook for this course is *Probability and Statistical Inference.*, by Robert Hogg, Elliot Tanis, and Dale Zimmerman. You may use either the latest (10th) edition or the previous (9th) edition.
You may also want to look at other textbooks to get a different perspective on the material:

- Hogg, R. V., Craig, A. and J. W. McKean, Introduction to Mathematical Statistics, 6th edition, Prentice Hall, 2004 (Higher level than Hogg and Tanis)
- McClave, J. T., Benson, P. G. and T. Sincich, Statistics for Business and Economics, 11th edition, Prentice Hall, 2010. (Lower level than Hogg and Tanis)

The readings for this course are about mathematical techniques and the "big picture" ideas that underly them. They are not bed-time reading. Take your time to prepare the readings for each class, and make sure you understand what is being presented. Preparation for class means doing some of the review exercises at the end of the assigned readings.

Software

Some of the course work, especially in the latter half of the course, will involve analysis of data using the Stata software package. There are three options for using Stata:

1. You can use Stata directly on your own computer by purchasing a six-month (or longer) license at http://www.stata.com/order/new/edu/gradplans/student-pricing/. I highly recommed the 6 month license for Stata/IC for $48.
2. You can also use Stata for free through a service called Apps on Demand that is accessible from the course site. After entering Apps on Demand, click on the Stata icon and you'll be running Stata inside your web browser window. You can connect this Stata environment to your Cornell Google Drive so you have persistent storage. It's not as good as running the software locally, but it works surprisingly well. The TA's will walk you through your initial set up during discussion section.
3. You may be able to use Stata in the public computer labs in Warren Hall and Mann Library (see http://www.cscu.cornell.edu/software/facilities.php).

I will spend some time in class teaching Stata and the program documentation is excellent. You will get plenty of practice during your sections, and there are several terrific free online resources for learning the software. For those students who feel more comfortable with a book in hand, Acock's *A Gentle Introduction to Stata, Sixth Edition* is up to date and pretty good.

Note to Students with Disabilities

If you have a disability-related need for reasonable academic adjustments in this course, please give me (Prof McKee) an accommodation letter from Student Disability Services. I expect you to give two weeks notice of the need for accommodations. If you need immediate accommodations, please arrange to meet with me within the first two class meetings.

Acknowledgements

Much of this class is derived from an econometrics class that Professor Lanier Benkard taught at Yale in Fall 2010. I'm extremely grateful to him for sharing his syllabus, lecture slides, assignments, handouts, exams, and advice. In addition, Francesca Molinari, Yongmiao Hong, and Jim Berry have generously shared materials that they have used in teaching this course and ECON 3120 in the past. All of these have provided a fantastic starting point. That said, I take full responsibility for any mistakes that I may have added to the material.

Schedule

PART I: PROBABILITY AND UNIVARIATE DISTRIBUTIONS

Module 1: The Big Picture

Lecture: August 22

Topics:
- Course overview

Module 2: Probabilities, Events, and Expectations

Lecture: August 24

Topics:
- Terminology and concepts: experiments, outcomes, and events
- Probabilities and chance
- Multiple events, probability rules, and Venn diagrams
- Probability tables
- Conditional probability: definition and intuition
- Probability trees

Module 3: Discrete Random Variables

Lecture: August 29 and 31

Topics:
- Probability mass functions
- Mean, variance, standard deviation, skew, and kurtosis
- Moment-generating functions
- Bernoulli and Binomial random variables

Module 4: Continuous Random Variables

Lecture: September 7 and 12

Due: Problem Set 1 on September 12, 8:00am

Topics:
- Probability density functions and cumulative distribution functions
- Uniform random variables
- Calculating Normal probabilities
- Sums of Normal random variables and standardizing

Module 5: More Random Variables

Lecture: September 14 and 19

Topics:
- Negative Binomial and Poisson distributions
- Exponential, Gamma, and Chi-Square distributions
- t and F distributions

Module 6: Review

Lecture: September 26

Due: Problem Set 2 on September 21, 8:00am

Thursday, September 29, 7:30pm FIRST PRELIM EXAM

PART II: MULTIVARIATE DISTRIBUTIONS AND ESTIMATION

Module 7: Bivariate Distributions

Lecture: September 21 and October 3

Topics:
- Bivariate discrete and continuous distributions
- Correlation and covariance
- Bivariate normal distribution

Module 8: Functions of Random Variables

Lecture: October 5 and 12

Due: Problem Set 3 on October 12, 8:00am

Topics:
- Functions of one random variable
- Functions of multiple random variables
- Markov's Inequality and Chebyshev's Inequality

Module 9: The Central Limit Theorem

Lecture: October 17

Topics:
- Intuition behind the CLT
- Formal proof of the CLT
- Application of the CLT

Module 10: Review

Lecture: October 19

Due: Problem Set 4 on October 19, 8:00am

Tuesday, October 25, 7:30pm SECOND PRELIM EXAM

PART III: ESTIMATION AND HYPOTHESIS TESTING

Module 11: Sample Statistics and Estimation

Lecture: October 26 and 31

Topics:
- Samples and populations: Overview
- Estimating a population mean and the law of averages
- Estimating and using variances, covariances, and correlations
- Standard errors
- Maximum Likelihood Estimation

Module 12: Sampling and Confidence Intervals

Lecture: November 2 and 7

Topics:
- Sampling and surveys
- Confidence intervals for means
- Confidence intervals for differences of two means
- Confidence intervals for proportions
- Sample Size

Module 13: Hypothesis Testing

Lecture: November 9 and 14

Topics:
- Null hypotheses and alternatives
- p-value mechanics for means and proportions
- Interpretation and communication of results
- One-sided vs. two-sided tests
- Small sample situations and t-tests
- tests of differences in means

Module 14: Nonparametric Tests

Lecture: November 16 and 21

Due: Problem Set 5 on November 16, 8:00am

Topics:
- Wilcoxon tests
- Chi-Square goodness-of-fit tests
- Assessing Normality

Module 15: Randomized Experiments

Lecture: Nov 28

Topics:
- Interpreting observational and experimental data

Module 16: Statistical Power

Lecture: Nov 30

Due: Problem Set 6 on November 30, 8:00am

Topics:
- A conceptual framework
- Power of a t-test
- Power functions
- Determining required sample sizes

Module 17: Review

Lecture: December 5