

ECON 3130: Introduction to Probability and Statistics

Cornell University, Fall 2017

Mon/Wed, 8:40am–9:55pm

Goldwin Smith Hall G76-Lewis

Instructor

Douglas McKee (douglas.mckee@cornell.edu)

Office: 466 Uris Hall

Office Hours: Wednesday 12:00pm–1:30pm

Teaching Assistants

Yu Leo She (ys693@cornell.edu)

Section 201: Friday, 9:05am–9:55pm, Warren Hall 150

Section 202: Friday, 10:10am–11:00am, Warren Hall 150

Office Hours: Tuesday, 1:00pm–2:30pm and Thursday, 3:30pm–5:00pm, Uris 465

Chris Rojas (car332@cornell.edu)

Section 203: Friday 1:25pm–2:15pm, Baker Lab 219

Section 204: Friday 2:30pm–3:20pm, Baker Lab 219

Office Hours: Tuesday, 3:30pm–5:00pm and Thursday, 10:00am–11:30am, Uris 451

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 introductory economics (ECON 1110 or 1120) and familiarity with single variable calculus (MATH 1110 and 1120).

Lectures

Lectures will be highly interactive. I will never talk continuously for more than 10 minutes at a time. You will be asked questions about the material and work on problems (alone and in small groups) several times per class. You will be using the i-clicker system during class to give me feedback on what you're learning and what you're not so I can tailor the lecture. This will require that you purchase either **an i-clicker hardware remote** or **the i-clicker REEF software for your mobile device**. The clicker costs \$30-40 and can be used for multiple classes. The software costs \$14.99 for a 6 month license. You can learn more on the Cornell Clicker How-to page: <http://www.it.cornell.edu/services/polling/howto-students.cfm>

Grades

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

1. Problem Sets (25%)

There will be 6 problem sets during the semester. You may pass in the problem sets as pdf's through the course web site by 8:40am or in person at the beginning of class on the due date. 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, including those turned in after class on the due date. Computer exercises must include the Stata .do and .log files that you produce.**

2. First Prelim (25%)

Date: Wednesday, September 27 (in class)

3. Second Prelim (25%)

Date: Wednesday, October 25 (in class)

4. Final Exam (25%)

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 September.

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.

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 A-B-C at proportions of 35-35-20 percent. The remaining 10 percent is spread across the letter grade distribution (including D and F) at the discretion of the instructor.

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 (<http://theuniversityfaculty.cornell.edu/handbook/Chapter5.pdf>) lists four types of situations in which faculty are encouraged to make accommodations for missed work. However, the determination as to whether a particular case warrants accommodation 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 not explicitly cumulative, but you will need to use concepts from the earlier parts of the course in order to understand the later topics. The tests may also include sub-questions that relate to earlier parts of the course if they are relevant.
- **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.

Acceptable Use Policy

You are free to use any published materials (e.g., another textbook), in preparing Econ assignments or for learning the material more generally. You are also strongly encouraged to work with others in your class. This is particularly helpful for learning to program. Each person must turn in their own assignment.

The use of any solution materials prepared in a previous year for ECON 3130, other than materials distributed this academic year by the course faculty, is strictly prohibited. This includes 1) any notes or handouts distributed by an instructor in a prior term of ECON 3130; and 2) any notes or solutions prepared by former students of Econ 3130, in either written or electronic form.

This policy means you should not solicit or use solutions to previous years' problem sets. The reason for this policy is that access to previous years materials can create serious inequities between fellow students, and jeopardize the integrity of the academic environment. Academic disciplinary actions will be taken against those who violate this policy.

Text and Readings

The required textbook for this course is *Probability and Statistical Inference.*, by Robert Hogg, Elliot Tanis, and most recently Dale Zimmerman. You may use either the latest (9th) edition or the previous (8th) edition which can be found used on Amazon starting around \$25.

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 recommend the 6 month license for Stata/IC for \$45. This is a terrific deal.
2. You can also use Stata for free by using the CISER account that will be created for all students in the class. This requires using a remote desktop connection to a CISER server. The performance is mediocre and transferring files back and forth is tedious. But it's free and it works. The TAs will show you how to do this in the first section. All students enrolled in the class before August 23 will have a CISER account setup automatically. If you enroll in the class after August 23, you need to fill out a new account for at http://ciser.cornell.edu/computing/accts/RSCH_ReqForm.php
3. You can 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, Revised Third 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 (Doug) 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.

Please do not redistribute any of these materials without my permission.

Schedule

PART I: PROBABILITY AND UNIVARIATE DISTRIBUTIONS

Week 1: Introduction

Lecture: August 23

Topics: – Course overview

Week 2: Probabilities, Events, and Expectations

Lecture: August 28 and 30

Read: HTZ (Hogg, Tanis, and Zimmerman) Chapter 1

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

Week 3: Discrete Random Variables

Lecture: September 6

Read: HTZ 9th ed. 2.1–2.4 or 8th ed. 2.1–2.5

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

Week 4: Continuous Random Variables

Lecture: September 11 and 13

Due: Problem Set 1 on September 13, 8:40am

Read: HTZ 9th ed. 3.1, 3.3 or 8th ed. 3.1, 3.3, 3.4, 3.6

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

Week 5: More Random Variables

Lecture: September 18 and 20

Read: HTZ 9th ed. 2.5, 2.6, 3.2 or 8th ed. 2.6, 3.5

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

Week 6: Prelim week

Lecture: September 25

Due: Problem Set 2 on September 25, 8:40am

Wednesday, September 27: FIRST PRELIM

PART II: MULTIVARIATE DISTRIBUTIONS AND ESTIMATION

Week 7: Bivariate Distributions

Lecture: October 2 and 4

Read: HTZ Chapter 4

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

Week 8: Functions of Random Variables

Lecture: October 11

Read: HTZ 9th ed. 5.1–5.3, 5.8 or 8th ed. 5.1–5.3, 10.5

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

Week 9: The Central Limit Theorem

Lecture: October 16 and 18

Read: HTZ 9th ed. 5.4, 5.6, 5.7, 5.9 or 8th ed. 5.4, 5.6, 5.7, 10.6

Due: Problem Set 3 on October 16, 8:40am

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

Week 10: Prelim Week

Lecture: October 23

Due: Problem Set 4 on October 23, 8:40am

Thursday, October 25: SECOND PRELIM EXAM

PART III: ESTIMATION AND HYPOTHESIS TESTING

Week 11: Sample Statistics and Estimation

Lecture: October 30 and November 1

Read: HTZ 9th ed. 6.1, 6.4 or 8th ed. 6.1

- 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

Week 12: Sampling and Confidence Intervals

Lecture: November 6 and 8

Read: HTZ 9th ed. 7.1–7.4 or 8th ed. 6.2, 6.3, 6.5, 6.6

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

Week 13: Hypothesis Testing

Lecture: November 13 and 15

Read: HTZ 9th ed. 8.1–8.3 or 8th ed. 7.1–7.4

- 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

Week 14: Nonparametric Tests

Lecture: November 20

Due: Problem Set 5 on November 20, 8:40am

Read: HTZ 9th ed. 9.1, 9.2, 8.4 or 8th ed. 8.1, 8.2, 8.5

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

Week 15: Randomized Experiments

Lecture: November 27 and 29

Read: Angrist and Pischke, Chapters 1 and 2 of *Mostly Harmless Econometrics* posted on web site

Due: Problem Set 6 on December 1, 5:00pm

Topics: – Interpreting observational and experimental data