

Econometrics and Data Analysis I

Yale University

ECON 131a, Fall 2015

Instructor: Doug McKee (douglas.mckee@yale.edu)

Lecture: MW, 2:30pm–3:45pm, Davies Auditorium

Teaching Fellows

Michael Puempel (michael.puempel@yale.edu)

Sections: TBD

Gerardo Ruiz Sancehez (gerardo.ruizsanchez@yale.edu)

Sections: TBD

Masayuki Sawada (masayuki.sawada@yale.edu)

Sections: TBD

Jeff Weaver (jeffrey.weaver@yale.edu)

Sections: TBD

Course Description

This course will teach you how to evaluate quantitative information and how to use data to answer quantitative questions in the social sciences. We will cover three main areas:

1. The first area, probability, provides a foundation for modeling uncertainties, such as the uncertainties faced by financial investors, insurers, and individuals in everyday life. We will study the mechanics of probability (manipulating some probabilities to get others) and the use of probability to make judgments about uncertain events.
2. The second area, statistics, provides techniques for interpreting data, such as what a marketing department might know about past consumer purchases. Statistical methods permit us to use small amounts of information to answer larger questions.
3. The third area, linear regression, is an area of statistics dedicated to estimating the relationships between two or more variables. For example, we might be interested in estimating the demand for health insurance as a function of the price, insurance policy attributes and consumer attributes.

The prerequisites for this course are introductory microeconomics and familiarity with single variable calculus.

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. By the end of the class, you will have acquired several concrete skills. Specifically, you will:

1. Understand the strengths and weaknesses of different methods.
2. Be able to choose appropriate methods to answer real-world questions.
3. Understand the math behind methods like linear regression and hypothesis testing.
4. Understand the intuition behind these methods.
5. Be able to apply these methods to analyze real data with a powerful statistical analysis package (Stata)

Grades

Your grade will be composed of five parts:

1. Problem Sets (20%)

There will be 7 problem sets during the semester. Problem sets 1-3 and 5-7 may be passed in through the course web site by 2:30pm or in person at the beginning of class on the due date. You do not need to pass in Problem Set 4. The teaching assistants will grade one randomly selected problem from each assignment (except for PS 4) and post complete solutions right 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.

2. Weekly Online Quizzes (10%)

Every Thursday, you will take an online quiz on that week's material. The questions are mostly easy and straight-forward and are designed to make sure you are paying attention to the lectures and reading. The quizzes are graded immediately and you will know what questions you got wrong. You can retake the quiz one time and your highest score will count as your grade. Annotated solutions to the quizzes will be posted every Friday morning. Your lowest quiz grade will be dropped.

3. Empirical Project (20%)

One of the most important parts of the class is the empirical project: It is your opportunity to use the tools you learn to answer a question **you** come up with and that **you** care about. At the end of the second week of classes, when the roster has settled down, I will divide the class into groups of four or five students each. Each group will identify and pass in their research question by the end of the third week (September 18). By the end of the 8th week (October 30), each group will pass in a document that quantitatively describes their sample and the relevant variables in it. During the last week of the classes (December 9), each group will do a poster presentation of their results, and during reading week (by December 14), groups will submit a written project report that includes a discussion of the empirical analysis. The research question is worth 5% of your final grade, the data description another 5%, and the

final analysis (presentation and report) is worth 10%. You will get a fair bit of written and verbal guidance throughout the process.

4. Midterm Exam (20%)

Date: October 19 (in class)

5. Final Exam (30%)

Date: TBD

Do not turn in anything late. Missed assignments or exams with no valid excuse receive a grade of zero. If you need to miss an exam or assignment any valid reason, please obtain a note from your residential dean and email me before the assignment is due or the exam date. Otherwise, you will get a grade of zero.

Exams are closed book, but you may bring one double-sided page of notes. You may use non-programmable calculators during the exams, and if you don't have one there will be simple calculators you can borrow during the exams.

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 should check out a “clicker” (looks like a cheap calculator) from the Yale Cross Campus Library before the first class. These are free, but you'll have to pay \$40 if you lose it during the semester. Once you have your clicker you will need to register it on the course web site.

Sections

Discussion sections in this class may be quite different from what you are used to as your TF's will not be summarizing the week's lectures or going over problems in front of the class. Instead, each week you will be using the methods you learned in lecture to analyze real data and answer real research questions. Everyone should bring their laptop with Stata installed and ready to go. Your entire hour will be spent interacting with your computer and each other, with an expert (your TF) nearby to answer any questions. In this way, the experience is very different from working on a problem set where getting stuck on something small for hours is common. Struggling with a problem is good for learning, but banging your head against a wall isn't always an efficient use of time. In addition, you will work in pairs and take turns driving—This will keep you focused and learning from each other.

Acceptable Use Policy

You are free to use any published materials (e.g., another textbook), in preparing Econ 131 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 131, other than materials distributed this academic year by the course faculty, is strictly prohibited. This includes 1) any notes, spreadsheets, or handouts distributed by me in a prior term of Econ 131; and 2) any notes, solutions, or spreadsheets prepared by former students of Econ 131, 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 *Introduction to Econometrics, 2nd or 3rd edition*, by Stock and Watson (Addison-Wesley, 2010). It's coverage of probability and statistics is somewhat rudimentary, but I will supplement this with other readings posted on the course web site. It's treatment of regression methods is excellent, and the book should serve you well as a reference in the future. Purchasing this book new is expensive, but you can typically find used copies of the textbook on the web at sites like Amazon for \$25.

For students without a strong mathematical background, you may also find the following (optional) text useful: *Probability and Statistical Inference, 8th or 9th ed.*, by Robert Hogg, Elliot Tanis, and most recently Dale Zimmerman (Pearson, 2010 or 2013). Hogg et al provides much deeper coverage of the concepts covered in the first half of the course than does Stock and Watson. The most important method we will cover during the course is linear regression and I highly recommend Paul Allison's *Multiple Regression: A Primer*. The writing is extremely clear and he covers both the intuition and mathematics behind the method.

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

Much of the course work in Econ 131, especially in the latter half of the course, will involve analysis of data using the Stata software package. Unless you already have access to Stata software, you must purchase at least a 6 month license for Stata/IC version 13. The student price is currently \$75 and it will allow you to do all the necessary analysis (and a whole lot more) on your own computer. To order Stata, contact Yale ITS or go to <http://www.stata.com/order/new/edu/gradplans/>

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.

You should try to attend an introduction to Stata class at the CSSSI at the beginning of the term. I will let you know when these classes are as soon as they are scheduled.

Additional Resources

A number of additional reading materials may also be of use. These include Anderson, Sweeney, and Williams, *Statistics for Business and Economics 6th ed.* (West Publishing, 1996), which provides a more mathematical treatment of the same material in the text; *Chance Encounters*, by Wild and Seber; Lapin, *Business Statistics* (Harcourt Brace, 1984), an excellent source of (solved) practice problems; *Decision Making Under Uncertainty* by Charles A. Holloway (Prentice-Hall, 1979), which provides an extensive treatment of formal decision analysis tools and techniques; and *The Visual Display of Quantitative Information* by Edward R. Tufte (Graphics Press, 1983), a masterful reference on the use and abuse of graphics and the visual display of data.

Acknowledgements

This class is in large part derived from the 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. 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 and Professor Benkard's permission.

Schedule

PART I: PROBABILITY

Week 1: Probabilities and Events

Lecture: September 2 and 4

Read: SW (Stock and Watson) Chapter 1
(Optional) HTZ (Hogg, Tanis, and Zimmerman) Chapter 1

Topics:

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

Week 2: Random Variables

Lecture: September 9

Read: SW 2.1–2.3
(Optional) HTZ 2.1–2.3

Topics:

- Probability trees
- Independence and information sets
- Terminology of random variables
- Expected values
- Variance and standard deviation
- Covariance and independence

Week 3: Binomial and Normal Random Variables

Lecture: September 14 and 16

Read: SW 2.4
(Optional) HTZ 9th ed: 2.4, 3.1-3.3; 8th ed: 2.4, 3.1-3.4, 3.6

Due: Problem Set 1 on September 14, 2:30pm
Research Question for Empirical Project on September 18

Topics:

- Calculating Binomial probabilities
- Expected Value and variance of Binomial random variables
- Calculating Normal probabilities
- Sums of Normal random variables and standardizing

PART II: STATISTICS

Week 4: Sample Statistics and Estimation

Lecture: September 21 and 23

Read: SW 2.5–2.6, 3.1

(Optional) HTZ 9th ed: 6.1, 7.1-7.4; 8th ed: 6.1-6.6

Topics:

- Sample and populations: Overview
- Estimating a population mean and the law of averages
- Estimating and using variances, covariances, and correlations
- Normal approximations and the Central Limit Theorem
- Distribution of sample sums and sample means
- Standard errors

Week 5: Sampling and Confidence Intervals

Lecture: September 28 and 30

Read: SW 3.3

Due: Problem Set 2 on September 28

Topics:

- Sampling and surveys
- Accuracy of sample statistics
- Confidence intervals for sample estimates
- Confidence intervals for averages

Week 6: Hypothesis Testing

Lecture: October 5 and October 7

Read: SW 3.2, 3.4, 3.6

(Optional) HTZ 9th ed: Chapter 8; 8th ed: Chapter 7

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
- Differences in means tests

Week 7: Randomized Experiments

Lecture: October 12 and 14

Due: Problem Set 3 on October 12

Problem Set 4 on October 14 (ungraded practice problems for midterm)

Read: SW 3.5

Angrist and Pischke, Chapters 1–2 (online)

Topics:

- Interpreting observational and experimental data

Monday, October 19: MIDTERM EXAM

PART III: REGRESSION

Week 8: Introduction to Regression Analysis

Lecture: October 26 and **online by October 28**

Read: SW Chapter 4

(Optional) Allison, Chapters 1 and 5

Due: Data Description for Empirical Project on October 30

Topics:

- Mechanics of simple regression
- Correlation vs. slope
- Interpreting regression estimates
- Doing regression in Stata
- R^2 and goodness-of-fit

Week 9: Statistical Inference in Regression

Lecture: November 2 and 4

Read: SW Chapter 5

Topics:

- Hypothesis testing and statistical significance
- Confidence intervals
- Prediction

Week 10: Multiple Regression

Lecture: November 9 and 11

Read: SW Chapters 6-7
(Optional) Allison, Chapters 2 and 3

Due: Problem Set 5 on November 9

Topics:

- Mechanics of multiple regression
- Interpreting multiple regression results
- Controlling for categorical variables with sets of dummy variables
- Regression F-test
- Joint Tests
- Restricted and Unrestricted models
- Tests of linear restrictions in regression models

Week 11: Model Building in Practice

Lecture: November 16 and 18

Read: SW Chapters 8–9
(Optional) Allison, Chapters 7 and 8

Topics:

- Developing regression models
- Interpreting and using results
- Model building using real data
- Confounding variables
- Intervening variables

Week 12: Linear Probability Models and Instrumental Variables

Lecture: November 30 and December 2

Read: SW 11.1, Chapter 12

Due: Problem Set 6 on November 30

Topics:

- The linear probability model
- Estimating causal effects with instrumental variables
- Evaluating instrumental variables

Week 13: Difference-in-Differences

Lecture: December 7 and 9

Read: SW 13.1–13.4
Angrist and Pischke, Chapter 5 (online)

Due: Problem Set 7 on December 7

Poster Presentation for Empirical Project on December 9

Topics:

- Using diff-in-diff to estimate policy effects
- Implementing and extending diff-in-diff with regression

Reading Week

Due: Data Analysis Report for Empirical Project on December 14