

Econometrics and Data Analysis I

Yale University

ECON S131 (ONLINE) – Summer Session B, 2017

July 3 – August 4

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

Teaching Fellow: Ana Reynoso (ana.reynoso@yale.edu)

Classroom: Online

Class Hours: MWF, 10:00am–11:00am

Instructor Office Hours (online): MWF, 11:00pm–12:00pm

TF Office Hours (online): TBD

Course Description

This course teaches 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:

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

The class is divided into thirteen modules spread over five weeks. After doing the assigned reading for a module, students are required to watch about an hour's worth of short videos. These videos are a mix of recorded lectures and me explaining concepts and doing math with pen and paper.

Each module also has a live session that is reserved for discussing the material and working through problems. For certain modules I will also ask you to do some preliminary data analysis before the live sessions.

Grades

Your grade will be composed of five parts.

1. Problem Sets (20%)

There will be four weekly problem sets that should be submitted by email before the beginning of class on the dates specified.

2. Thirteen Online Quizzes (10%)

Before every live session, you will take an online quiz on that module'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. They are also designed to help me identify the topics I need to focus on during the live sessions. The quizzes are graded automatically and you will know immediately what questions you got wrong. You can retake the quiz one time and the highest of your two scores will count. Annotated solutions to the quizzes will be posted after every live session. Your lowest quiz grade will be dropped.

3. Empirical Project (20%)

On Tuesday of the first week I will divide up the class into 3-4 person groups. Each group will identify a research question by the beginning of the second week. By the end of the third week, each group will pass in the data they are planning to analyze along with a memo describing the relevant variables in the data. On Wednesday of the last week, each group's project report including a discussion of the empirical analysis is due. We will talk more about this assignment during class. The research question is worth 5% of your final grade, the data description another 5%, and the analysis itself is worth 10%.

4. Midterm Exam (20%)

Date: July 19

5. Final Exam (30%)

Date: August 4

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 \$50.

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 required 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 14. 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.

I will spend some time in class teaching Stata and the program documentation is excellent. In addition, there are several terrific free online resources. For those students who feel more comfortable with a book in hand, Acock's *A Gentle Introduction to Stata, Fifth Edition* is up to date and pretty good.

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

(M1) Monday, July 3: Probabilities and Events

Read: SW, 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

(M2) Wednesday, July 5: Random Variables

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

(M3) Friday, July 7: Binomial and Normal Random Variables

Due: Problem Set 1

Read: SW 2.4

Binomial distribution hand-out

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

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

(M4) Monday, July 10: Sample Statistics and Estimation

Due: Research question for empirical project

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
- Introduction to Stata

(M5) Wednesday, July 12: Sampling and Confidence Intervals

Read: SW 3.3

Topics:

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

(M6) Friday, July 14: Hypothesis Testing

Due: Problem Set 2

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

(M7) Monday, July 17: Randomized Experiments

Due: Problem Set 3

Read: SW 3.5

Angrist and Pischke, Chapters 1–2 (online)

Topics:

- Interpreting observational and experimental data

Wednesday, July 19: MIDTERM EXAM

PART III: REGRESSION

(M8) Friday, July 21: Introduction to Regression Analysis

Due: Data description for empirical project

Read: SW Chapter 4

Optional: Allison, Chapters 1 and 5 3

Topics:

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

(M9) Monday, July 24: Statistical Inference in Regression

Read: SW Chapter 5

Topics:

- Hypothesis testing and statistical significance
- Confidence intervals
- Prediction

(M10) Wednesday, July 26: Multiple Regression

Read: SW Chapter 6

Optional: Allison, Chapters 2 and 3

Topics:

- Mechanics of multiple regression
- Interpreting multiple regression results
- Confounding variables
- Intervening variables
- Dummy regression variables

- Model building and avoidable mistakes

(M11) Friday, July 28: Model Building in Practice

Read: SW Chapters 8–9

Optional: Allison, Chapters 7 and 8

Topics:

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

(M12) Monday, July 31: Joint Hypothesis Testing

Due: Problem Set 4

Read: SW Chapter 7

Topics:

- Joint hypothesis testing
- Regression F-test
- Restricted and Unrestricted models
- Tests of linear restrictions in regression models
- The linear probability model

(M13) Wednesday, August 2: Difference-in-Differences

Due: Final report for empirical project

Read: SW 13.1–13.4

Angrist and Pischke, Chapter 5 (online)

Topics:

- Natural experiments and regression
- Strategies for identifying effects of a policy
- Extending diff-in-diff with a regression

Friday, August 4: FINAL EXAM