Econometrics and Data Analysis I

Yale University
ECON S131 (ONLINE) – Summer Session B, 2013
July 8 – August 9

Instructor: Doug McKee (douglas.mckee@yale.edu)
Classroom: Online
Class Hours: MWF, 10:00am–11:00am
Office Hours (online): MWF 11:00am–12:00am

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 the data a marketing department might have on 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 or 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)
Students are required to watch about three hours of video content each week. These videos are a mix of recorded lectures and me explaining concepts and doing math with pen and paper. You should be doing the reading before spending time with the video.

Live class time will be reserved for discussing the material and working through problems. For certain classes I will also ask you to do some preliminary data analysis before the live sessions.

**Grades**

Your grade will be composed of four parts.

1. **Problem Sets** (20%)  
   There will be four weekly problem sets, distributed via the classes*v2 website, and due at the beginning of class on the dates specified.

2. **Empirical Project** (20%)  
   There will also be a (longer) empirical project on a topic and dataset to be determined by me that will be handed out during the term and that will be due in the last week of class.

3. **Midterm Exam** (25%)  
   Date: July 24

3. **Final Exam** (35%)  
   Date: August 9

**Text and Readings**

The required textbook for this course is *Complete Business Statistics, Eighth Edition*, by Amir Aczel (Wohl, 2012). This book is expensive. However, previous editions of the book are also fine, and you can typically obtain previous editions of the textbook on the web at sites like half.com for less than $10. You do not need the CD that comes with the book.

For students without a strong mathematical background, you may also find the following (optional) text useful: *Statistics, 4th ed.*, by David Freedman, Robert Pisani, and Roger Purves (Norton, 2007). Freedman et al. provides a lucid introduction to many important (and difficult) concepts in statistics, at a more accessible level than the required text. It has been well-received in the past by Econ 131 students for whom statistics was a new subject. The work horse of 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.

We will also employ some additional readings, notes and cases during the course. These will
be distributed on the course web site during the term. All three suggested textbooks should be available at the Yale Bookstore or from your favorite online bookseller.

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 13. The student price is currently $69 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, Revised Third 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

Monday, July 8: Probabilities and Events

Read: Aczel, 2.1–2.4
Optional: Freedman et al, Chapters 13–14

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

Wednesday, July 10: Random Variables

Read: Aczel, 2.5–2.8, 3.1–3.4 (3.1–3.3 in 7th edition)
Optional: Freedman et al, Chapter 14

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

Friday, July 12: Binomial and Normal Random Variables

Due: Problem Set 1

Read: Aczel, 3.5–3.6 (3.4–3.5 in 7th edition), 4.1–4.5
Optional: Freedman et al, Chapter 15

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

Monday, July 15: Sample Statistics and Estimation

Read: Aczel, 1.1–1.5, 4.7, 5.1–5.3, 6.1–6.2
Optional: Freedman et al., Chapter 20

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

Wednesday, July 17: Sampling and Confidence Intervals

Read: Aczel, 6.3–6.4
Optional: Freedman et al., Chapters 21 and 23

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

Friday, July 19: Hypothesis Testing

Due: Problem Set 2
Read: Aczel, 7.1–7.4, 8.1–8.4
Optional: Freedman et al., Chapter 27

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
Monday, July 22: Randomized Experiments

Due: Problem Set 3  
Read: Angrist and Pischke, Chapters 1–2 (online)  
Topics:
  – Interpreting observational and experimental data

Wednesday, July 24: MIDTERM EXAM (10:00–12:00)

PART III: REGRESSION

Friday, July 26: Introduction to Regression Analysis

Read: Aczel, 10.1–10.4, 10.6, 10.10  
Optional: Allison, Chapters 1 and 5  
Topics:
  – Mechanics of simple regression  
  – Correlation vs. slope  
  – Interpreting regression estimates  
  – Doing regression in Stata  
  – $R^2$ and goodness-of-fit

Monday, July 29: Statistical Inference in Regression

Read: Aczel, 10.9, 11.1, 11.2, 11.4  
Topics:
  – Hypothesis testing and statistical significance  
  – Confidence intervals  
  – Prediction

Wednesday, July 31: Multiple Regression

Read: Aczel, 11.1–11.9  
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

Friday, August 2: Model Building in Practice
Read: Aczel, 11-10–11.11
Optional: Allison, Chapters 7 and 8

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

**Monday, August 5: Joint Hypothesis Testing**

Read: Aczel, 11.13

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

Due: Problem Set 4

**Wednesday, August 7: Difference-in-Differences**

Read: 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

Due: Empirical Project

**Friday, August 9: FINAL EXAM (10:00–12:00)**