STA 371: Statistics and Modeling
Course Syllabus
Spring 2012

Course overview

THE PURPOSE OF statistics is to help answer tough questions in the face of uncertainty and randomness. What policies contribute the most to creating sustained economic growth? Do charter schools work miracles for their pupils, or benefit from self selection? Should your company become an early adopter of an expensive new technology whose ultimate payoff is in doubt? Questions like these can't be studied using controlled experiments. Rather, we must carefully sift through a body of evidence, hoping to tease out relationships in complex, noisy systems that don't always behave like they would if we could draw things up in a laboratory.

In this course, you will learn to use the language of probability to study problems such as these in a formal quantitative way. My goal is to help you cultivate two important skills:

1. Building simplified models of real-world systems to aid cause-and-effect reasoning and guide intelligent behavior.

2. Using visual and quantitative evidence to evaluate hypotheses in loosely structured problems with no verifiably correct answer.

Throughout the course, I will emphasize the analysis of real data, and will provide examples from finance, politics, sports, marketing, economics, and science to help illustrate the concepts you'll be learning. By semester's end, you will have learned some lessons that will serve you well throughout the rest of your life, both as a producer and as a consumer of statistical information.

The methodological focus of this course is on building statistical models for data analysis and decision making. We will be concerned mainly with understanding and interpretation, rather than theoretical details. This does not mean that we won't encounter some math along the way—just that the models themselves, rather than their formal properties, will be the focus.

The outline on the following page is subject to review if we need to slow down or speed up, but it should give you a more detailed idea of how the course will proceed.

Week	Topic	Day by day
1	Explanation and evidence	Course nuts and bolts; head start on Chapter 1 An introduction to data-based reasoning
2	Exploring multivariate data	Variation across categories; variation between, and within, groups Two quantities varying together; GDP growth case study
3	Predictable and unpredictable variation	Sums of squares in group-wise models and regression The analysis of variance; interpreting R^2
4–5	Quantifying uncertainty, part 1	Sample, resample: a tale of 10,000 maps Fishing rods, bootstraps, and the role of uncertainty in regression The role of probability in statistical modeling Estimation, testing, and prediction in one-variable regression
6	Beyond straight lines	Fitting non-linear curves Modeling binary outcomes
7	Multiple regression	Fun with charismatic megafauna: a first look at multiple regression Aggregration paradoxes and dummy variables
8	Midterm week	Review and catch-up Midterm exam
9–10	Quantifying uncertainty, part 2	Estimation, prediction, and testing in the multiple regression model Collinearity and the subtleties of an ANOVA table Nothing is automatic: on choosing a model "The smuggling of art, the art of smuggling": a case study in modeling
11	Time series	As spring follows winter: measuring trends and seasonalities When history rhymes: mean reversion and autoregressive models
12	Modeling randomness	Probability and simulation: a refresher Correlated random variables
13	Utility	Nicolas Bernoulli and the St. Petersburg paradox The art and science of the "bad trade"
14	Making smart decisions	How much to pay How much to gamble
15	Putting it all together	Case studies in modeling, simulation, and decision analysis

How the course is structured

On a day to day basis, this course revolves around independent inquiry, in addition to traditional lectures. The focus of class time is on building your capacity to think about open-ended problems, not conveying "material." On any given day, there will be a mix of lecture, discussion, and hands-on modeling.

The upshot of all this? You will end up learning many important skills outside of class. This learning will take (at least) three forms:

Reading. Mostly this will be out of the online course packet, but there will also be supplemental readings from other sources.

Watching video tutorials. Some of these are about statistical concepts, and some are about software.

Practice. Weekly problem sets will have two parts: (1) practice problems, which won't be graded, and are there for you to check your understanding of the basics; and (2) one or two problems to turn in. You are encouraged, but not required, to work on these in groups of four people or fewer.

You'll find links to all the relevant material through the class website. The Home tab keeps a blog-style diary of what readings and videos you should keep up with, what practice problems you should be working on, and so forth. The Data and Code, Exercises, and Videos tabs are all pretty self-explanatory. The Other Resources tab has links to relevant material contained on external websites. I will often refer to these materials when suggesting supplemental readings.

As you might imagine from this description, succeeding in this course will require substantial time devoted to out-of-class preparation. Will it be worth it? Absolutely! You will learn a lot about statistics if you make the effort. As a rule of thumb, you should expect to spend 3 hours per week in class; 1–2 hours per week reading and watching videos; and anywhere from 3-6 hours per week completing the exercises. As with all college classes, in some weeks there will be less than this, and in some weeks there will be more.

Materials and software

If you do not wish to, you don't have to spend a single penny on books or software for this course.

As far as books go: I hate the idea of assigning an expensive textbook that is, at best, a halfway match with the goals of the course. So I haven't done this. Instead, our main reference will be a free online course packet (*Statistical Modeling: A Gentle Introduction*). I wrote the packet myself, so it is carefully integrated with the rest of the course material. I will post chapters to Blackboard as they become relevant over the semester.

As for software, we will use an open-source statistical package called R (www.r-project.org) for almost of what we need to do. R is freely available for Mac, Windows, and Linux. It's the real deal—no silly Excel limitations for us. I want you to come away from the course with a legitimate, industrial-strength platform for hands-on data analysis. There's also a nice, free interface to R called R Studio: www.rstudio.org. You might like this better than the standard Mac or Windows interface.

Strictly speaking, R is recommended, not required: if you're an Excel whiz, you might be able to make it work for you for the entire course. If you are tempted to take this route, though, you should consider a few things: 1) R is easy to learn, and the course will hold your hand a lot in the beginning; 2) R is drastically more powerful, efficient, and versatile than Excel for data analysis; and 3) the plug-in you need to do statistical modeling in Excel isn't available for Macs. But you are 100% free to use any software package you like, including Matlab, STATA, SAS, Minitab, etc.

Prerequisites

The formal university prerequisites for this course are: Business Administration 324 or 324H; Management Information Systems 301 or 310; Mathematics 408D, 408L, or 408M; and Statistics 309 or 309H. Calculus is a pre-requisite, but you won't need it at a particularly advanced level. Here's a simple diagnostic: can you differentiate the function $f(x) = \ln x$ with respect to x? If you can, then you know enough calculus to succeed in this course. If you needed to look up the answer on Wikipedia to remind yourself, but you get the gist of it, then you're probably OK. If you don't understand what the question is asking, then you need a math

refresher before you take this course.

Exams and grading

Grades will be determined by two exams (20% each), two projects (20% each), and regular homework assignments (20% total).

Homework will count for 20% of your final grade. All homework must be turned in by the end of class on the day it is due. Homework is graded on a 10-point scale. No late homework will be accepted, but everybody gets two drops (basically two free 0's that won't count towards your grade).

The midterm is worth 20% of your final grade, and will take place on the last day of class before Spring Break. You will be allowed to bring a calculator, but it is not necessary to have one. The exam will be graded such that, if you set up all calculations in the appropriate way, you will get full credit even if it is not possible to get the final answer without a calculator.

If you must miss the exam for the observance of a religious holy day, inform me as far in advance of the day as possible, so that alternative arrangements can be made in conjunction with the Dean and the relevant university offices. If you miss the midterm for any other reason—including illness or travel—then you must inform me in advance, and I will allow you to count your final exam grade as your midterm grade. This option cannot be exercised retroactively, and it is not available unless you inform me in advance. (If you are sick, then an e-mail on the morning of the mid-term will be fine.)

There are two course projects, which you may complete in groups of four people or less if you wish. (Groups are optional.) Each project is worth 20% of your grade. The first project will involve getting your own data set on a question that interests you, running an appropriate statistical analysis, and writing up your conclusions. The details will be discussed in class, and posted on the course website. This project is due by Friday, March 23, 2012.

The second project will involve the use of a common data set to help you in a complex, real-world decision-making problem. Again, the details will be discussed in class and posted on the course website. This project is due by Friday, May 4, 2012.

A uniform, in-class final exam for both sections will take place during the usual University exam period in early May, and will count for 20% of your grade. I will announce the time slot as soon

Grading
Homework: 20%
Midterm: 20%
Final exam: 20%
Project 1: 20%
Project 2: 20%

Mid-term: in class on the last class day before spring break

Project 1: due Friday, March 23, 2012.

Project 2: due Friday, May 4, 2012.

Final
When: TBA, during offical exam period
Where: TBA
Make-up: TBA

as it is available from the University registrar. If you can't make the scheduled time because of a conflicting exam, there will be a different, make-up exam scheduled on a different day.

Re-grade requests

On occasion you may notice a simple clerical error in the recording of a grade, which I am happy to correct without hassle. Other regrading requests must be submitted in writing within 7 days of the marked paper being returned. Keep in mind that the entire paper will then be subject to re-grading, and that your grade may go up or down as a result.

Attendance

Beyond the obvious correlation between coming to class and overall course performance, attendance does not play a role in course grading.

Curving grades

The raw percentage scores to the right will guarantee you *at least* the corresponding grade.

I reserve the right to curve grades up. But I will never curve them down. That means these grades are a floor, not a ceiling, on the final grade that someone with the corresponding raw score would receive. The precise details of any curve are at my sole discretion, and if I should choose to use a curve, I will detail the cutoffs used when course grades are submitted.

Percentage	Grade
93–100	A
90–92	A-
87–89	B+
83–86	B
80–82	B-
70–79	C
60–69	D

Other course details

Quantitative reasoning flag

This course carries the Quantitative Reasoning flag. Quantitative Reasoning courses are designed to equip you with skills that are necessary for understanding the types of quantitative arguments you will regularly encounter in your adult and professional life. For more details, see `www.utexas.edu/ugs/core/flags/quantitative-reasoning`.

Classroom etiquette

You are expected to participate in class; close and put away your laptops, unless it's a "hands-on" day where I ask you to look at data and run models in class; and to turn off your phones, iPods, and other cool gizmos. I also ask that you arrive on time to class, since late arrivals disrupt things for all other students. In turn, I will make sure we finish on time so that students may reach their next lectures/hot dates.

Cheating, plagiarism, and such

Acts of academic dishonesty are ethically wrong; they harm the reputation of the school and demean the honest efforts of the majority of students. You know it; I know it; and no excuses will be accepted. Additionally, you should consider three things:

(1) Cheaters are a tiny minority. The vast majority of students who preceded you did it the honest way. Follow their lead.

(2) You play like you practice. The habits you form now will predict the headlines that people write about you, or your company, later in life. Try Googling "Jeff Skilling" or "Fabulous Fab" if you don't believe me.

(3) If you cheat, you're playing with fire. The minimum penalty will be a zero for that assignment or exam. You also risk failing the course and being dismissed from the University.

The bottom line when it comes to cheating is: just don't do it. You might fool me, if you're very lucky and very unscrupulous. But you are highly unlikely to fool the McKinsey interviewer you were hoping to impress with your knowledge of statistics. And you may find that the job market is far more ruthless than university judicial boards.

Now for the usual boilerplate. The responsibilities of both students and faculty with regard to scholastic dishonesty are described in detail in the Policy Statement on Scholastic Dishonesty for the McCombs School of Business. By enrolling in this class, you have agreed to observe all of the student responsibilities described in that document. By teaching this course, I have agreed to observe all of the faculty responsibilities described in that document.

My first hit for "Fabulous Fab" is the *New York Daily News* from 27 April 2010, which wrote: "Fabrice Tourre, who calls himself Fabulous Fab, is not so much. Actually, the 31-year-old Frenchman of the racy e-mails came across like a weenie when he appeared before a Senate subcommittee to be grilled about Goldman Sachs' role in a deal the SEC says wasn't kosher." Cheat at your own risk, weenie.

Students with disabilities

The University of Texas at Austin provides upon request appropriate academic accommodations for qualified students with disabilities. Services for Students with Disabilities (SSD) is housed in the Office of the Dean of Students, located on the fourth floor of the Student Services Building. Information on how to register, downloadable forms, including guidelines for documentation, accommodation request letters, and releases of information are available online at `deanofstudents.utexas.edu/ssd/index.php`. For more information, contact the Office of the Dean of Students at 471-6259, or 471-4641 TTY.

Student privacy

First of all, you should know that I am legally barred from discussing your course performance with anyone other than you and anyone that you explicitly designate. That includes your parents.

Second, a note on Blackboard. Blackboard is a password-protected web site, and is created automatically for all accredited courses taught at The University. I will post the syllabus, handouts, assignments and various other resources on Blackboard. Other site activities could include exchanging e-mail, engaging in class discussions and chats, and exchanging files. In addition, Blackboard include a class e-mail roster. Students who do not want their names included in such an electronic class rosters must restrict their directory information in the Office of the Registrar, Main Building, Room 1. For information on restricting directory information, see `www.utexas.edu/student/registrar/catalogs/gi02-03/app/appc09.html`.