**ECO s394D: Probability and Statistics**

Instructor: James Scott (james.scott@mccombs.utexas.edu)

Lectures: M T W Th, 1:30 PM to 3:15 PM, **BRB 1.118**
TA sessions: Fridays 1:30 PM to 2:30 PM, also **BRB 1.118**

Office hours:
M T W 12:30 PM to 1:20 PM (**CBA 6.478**)

Teaching Assistant: Nate Hattersley (nhattersley@utexas.edu)
Office hours TBA

**Background and objectives.** This course is an intermediate level introduction to probability theory and statistics. The main goals are: 1) to give you some practical tools for data analysis in R; and 2) to prepare you for econometrics. Students taking the course should have a working knowledge of calculus and basic descriptive statistics. The emphasis will be on conveying the basic principles of probability, data analysis, and statistical theory. This is fast-paced course with a condensed schedule, so it is essential to attend lectures regularly and keep up with the material.

**Course materials.** There's nothing to buy. Your main points of reference will be:

- lecture notes from class, typically provided as slides via Canvas**.**
- "Data Science in R: A Gentle Introduction" I will refer to this as "DSGI" in the syllabus.
- "Introduction to Probability Lecture Notes," by Dimitri P. Bertsekas and John N. Tsitsiklis. This is posted as a PDF file on the course Canvas page; in the syllabus, I will refer to it as "Bertsekas".
- Statistical Inference for Data Science, by Brian Caffo. This is free to read online, or you can buy a copy on a "name your own price" model (where $0 is a valid price). I will refer to this as "Little Inference" in the syllabus.

Since the material we're covering is pretty standard, basically any intermediate probability/statistics textbook can be used as a supplement.

**Lectures.** Lectures will take place in person in BRB 1.118. There is no remote participation option, but lectures will be recorded and available to watch after class via Canvas.

**Communication.** I will post all assignments, announcements, and class materials on our class's Canvas site.

**Software.** We will R and RStudio for data analysis. For details on how to install them, see the first chapter of "Data Science in R: A Gentle Introduction."

**Assignments and grading: 70% homework, 30% final exam**.

There are four weekly problem sets, each graded as a percentage out of 100. The

weekly problem sets are posted on Canvas and you will submit your assignments electronically through Canvas. Each assignment should be submitted as a single file or as a link. All files must be in PDF format (no other formats accepted).

You may discuss the problem sets with classmates, but you may not copy each other's work. Each student must submit their own write-up for each homework.

In the past, students have found it pretty simple to just take pictures of their written work with a phone, and then stitch the individual pictures together into a single file using any appropriate piece of software (e.g. Mac Preview, Microsoft Word, Google docs. etc). You can also write your solutions and post them at a shareable link, e.g. a Google doc (including images, e.g. phone pictures of handwritten work) and submit that link via Canvas. If you do this, make sure you change the permissions so that anyone with the link can view the document.

In grading the homework, we will be looking at the following:

- Did you make an honest, concerted attempt at each problem?
- For a math problem: did you show all your steps? If you were stuck, did you show evidence of attempting to get yourself unstuck, e.g. by writing down results you know that you think might be relevant?
- For a data-analysis problem: did you attempt to address all parts of the question? Did you include figures/tables where appropriate? Did you write up your solution with some semblance of narrative flow and professionalism (good), or did you just copy and paste a bunch of R code without much in the way of explanation (bad)?
- And finally, of course, did you make any significant technical mistakes?

The Friday TA session will be geared towards helping students with that week's problem set, which will be due at the beginning of class the following Monday. Late homework turned in within 24 hours of the original deadline will receive a 20% deduction (i.e. a maximum score of 80). Homework turned in after this will receive a grade of 0.

At the end of a semester, we will average your four HW scores, and this HW average will count for 70% of your final grade.

There is a cumulative final exam worth 30% of the course grade, which will take place on the day assigned by the university registrar's office. Specific date and time TBA.

**Course topics outline**

*Getting started with data.* Data frames, cases, and variables. Samples vs. populations. The unit of analysis. Getting started with R. (DSGI Lessons 1-2).

*Counting + basic probability.* Conditional and joint probabilities. Negation, addition, and multiplication rules. Probabilities estimated from data. Independence; checking independence from data. (DSGI Lesson 3).

*Plots.* Five key plots. Best practices for data visualization (DSGI Lesson 4)

*Data summaries and data wrangling* (DSGI Lessons 5-6)

*Fitting equations to data* (DSGI Lesson 7)

*Further probability topics.* Random experiment. Set operations, Kolmogorov's axioms. Interpretations of probability. Basic rules/probability calculus. Joint, conditional, and marginal probability. Simpson's paradox. Rule of total probability. Bayes' rule. (Bertesekas Ch 1)

*Discrete random variables and their distributions.* PMFs: binomial, Poisson. Expectation and variance. (DSGI Lesson 17.1-17.2, Bertsekas 2.1-2.2, 2.4; supplemental reference in Little Inference Chs 2, 4, 5, 6.).

*Continuous random variables and the CDF/PDF.* Uniform, exponential, and normal distributions. (DSGI Lesson 17.3, Bertsekas 2.3, 3.1-3.3, 3.6; supplemental reference in Little Inference Ch 6.)

*Multivariate distributions.* Joint, marginal, and conditional distributions, covariance and correlation; linear combinations of random variables. If time: the bivariate normal distribution (Bertsekas 2.5-2.7, 3.4-3.5, 4.5, 4.7)

*Statistical uncertainty.* Standard errors; bootstrap; confidence intervals (DSGI Lessons 8-9)

*p-values* (DSGI Lesson 10). Null hypothesis, test statistic, sampling distribution, p-value. Interpreting p-values.

*Large-sample inference.* Convergence in probability. The Law of Large Numbers. Consistency. Convergence in distribution. The Central Limit Theorem. Inference based on asymptotic normality. The t-distribution. (DSGI Lesson 11, Bertsekas Ch 7, Little Inference Ch 7.)

*Testing in parametric models.* Null and alternative hypotheses, confusion matrix, Type 1 and 2 errors. Alpha level vs. p-values. (Little Inference Ch 9, 10)

*Constructing estimators.* Maximum likelihood, method of moments. (See constructing_estimators.pdf on course Canvas page under Files tab).

**University policies:**

Honor code: "The core values of the University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the University is expected to uphold these values through integrity, honesty, trust, fairness, and respect towards peers and community."

Any student with a documented disability (physical or cognitive) who requires academic accommodations should contact the Services for Students with Disabilities area of the Office of the Dean of Students at 471-6259 (voice) or 471-4641 (TTY for users who are deaf or hard of hearing) as soon as possible to request an official letter outlining authorized accommodations.

Students who violate University rules on scholastic honesty are subject to disciplinary penalties, including the possibility of failure in the course and dismissal from the University. Since dishonesty harms the individual, fellow students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced.

Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside. Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building. In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office. Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.

Behavior Concerns Advice (BCAL): 512-232-5050.

*Harassment Reporting Requirements.* Senate Bill 212 (SB 212), which went into effect as of January 1, 2020, is a Texas State Law that requires all employees (both faculty and staff) at a public or private post-secondary institution to promptly report any knowledge of any incidents of sexual assault, sexual harassment, dating violence, or stalking "committed by or against a person who was a student enrolled at or an employee of the institution at the time of the incident". Please note that both the instructor and the TA for this class are classified by SB 212 as mandatory reporters. That means we MUST share with the Title IX office any information about sexual harassment/assault that is shared with us by a student—whether in-person, via electronic communication, or as part of any class assignment. Note that a report to the Title IX office does not obligate a victim to take any action, but this type of information CANNOT be kept strictly confidential except when shared with designated "confidential employees." A confidential employee is someone a student can go to and talk about a Title IX matter without triggering any obligation by that employee to have to report the situation so that it will be investigated. A list of confidential employees is available on the Title IX website. The professor and TA for this class are NOT designated confidential employees per SB 212.