Syllabus for Econ 20, Spring 2021

Instructor: Isaiah Andrews
Office Hours: Wednesday 2:00-4:00 PM
Email: iandrews@fas.harvard.edu
Teaching Fellow: Francois-Xavier (FX) Landant
Office Hours: TBD
Section: TBD
Email: fladant@g.harvard.edu

Course Overview: This course will introduce students to data analytic methods useful for answering social science questions. The course will cover the fundamentals of probability and statistics while introducing students to causal inference, quasi-experimental methods, and regression analysis. All the methods studied in the course will be motivated and illustrated with real-world applications. This course does not require any prior coursework in economics, but is complementary to Econ 50 (Using Big Data to Solve Economic and Social Problems) and satisfies the prerequisites for Econ 1123 (Introduction to Econometrics).

For economics concentrators, Econ 20 counts as a secondary course and counts towards the writing requirement. It does not count as an elective.

Course Meetings: The course will combine pre-recorded lectures and small-group meetings. Lectures will be posted on Canvas (under "cloud recordings" in the Zoom tab), with several videos per lecture. In addition, all students should sign up for a half-hour small-group meeting slot. The goal of these small-group meetings is to discuss the course material and answer questions. Please attend the same group meeting slot every week, watch the lectures up to and including the one for that day, and come

with questions.

We have provided six initial group meeting slots, based on the scheduled course meeting times, but will add more if needed. If you cannot make the scheduled course time due to time zone issues, please let us know on the sign-up sheet, and we will reach out to schedule additional meeting times. **Sign up here:** https://bit.ly/3rTZxBt

Enrollment: There is currently an enrollment cap or 40 for this course. This cap is higher than the the total course enrollment last year, so I expect to be able to enroll everyone who is interested. Moreover, in the event there are more than 40 students interested in the course, I also expect to be able to increase the cap to some degree (limited by the need to keep the group meetings a reasonable size). In the unlikely event we are not able to accommodate everyone in the course, I'll hold a lottery on 1/19, and approve course enrollments for the largest number of people possible.

Introductory Materials: Videos for the first lecture, giving an overview of the course, are posted on Canvas. For any questions in advance of the start of term, I will hold office hours at the regular time (Wednesday, 2:00-4:00pm) on 1/13 and 1/20.

Grading: Grading will be based on five empirical exercises, six online multiple-choice problem sets, a midterm exam, and a final exam. The course grade will assign weight 10% to each of five empirical exercises, 10% (total) to the five problem sets, equally weighted, 10% to attendance at the small-group meetings, 15% to the midterm exam, and 15% to the in-class final exam.

Examinations: There will be an open-book take-home midterm for the class on Tuesday 3/9, and a take-home final on Thursday 5/6.

Assignments: There will be two types of assignments in the course: problem sets, and graded empirical exercises.

Problem Sets: Problem sets will be posted every two weeks. The purpose of the problem sets is to help you master the material covered in class, and to practice skills for the empirical exercises and exam. The problem sets will be multiple choice. You may discuss the problem sets with your classmates, but we strongly encourage

you to at least attempt all the questions on your own first, as this is much better practice for the exams. Problem sets are due **at the beginning of the scheduled course time (10:30am ET)** on the days listed below.

Empirical Exercises: These exercises will ask you to analyze some features of a dataset, precisely describe the methods you use and the results you obtain, and interpret those results. Some further details:

- Your writeup for each empirical exercise should be three or four pages (1.5 spaced) and typewritten.
- Each empirical exercise must be accompanied by code that produces the results (i.e. that takes in the raw data and outputs any numbers or figures that appear in your writeup).
- We ask that code be written in the statistical package R, which is available for free at https://www.r-project.org/. For those getting started with R we recommend using it through R Studio, which is likewise available at https://rstudio.com/ and should be installed after R.
- **You are welcome to discuss the empirical exercises with classmates, but should write both your code and solutions for yourself.** Identical write-ups (or code) will receive no credit. Please list everyone with whom you discussed or collaborated at the end of your writeup.
- Empirical exercises are due **by the beginning the scheduled class time (10:30am ET)** on the day listed.
- Please submit empirical exercises (and code) via Canvas
- Late empirical exercises will be accepted, but 20 percentage points will be deducted from the grade for each day late. Hence, exercises late by 5 days or more will receive zero credit.

Prerequisites: There are no prerequisites for the course. High school-level algebra will be assumed. The course is suitable for students whether or not they have taken AP statistics.

Readings: The course material is self-contained and there is no textbook. Slides covering the material will be presented in class and posted on the website. However, some students may be interested in additional references. Some good references for the material in the course are:

Stock, J. and M. Watson (2019), *Introduction to Econometrics* (4th Edition). Pearson, New York.
(First four chapters cover material relevant to this course. Textbook for Econ 1123)

Wheelan, C. (2013), *Naked Statistics*. W. W. Norton, New York.
(Informal discussion of many of the probability and statistics topics covered in this class, with many examples)

Angrist, J. and S. Pischke (2015), *Mastering Metrics*. Princeton University Press, Princeton.
(Discusses approaches to uncovering causality, again with many examples)

Additional references will also be flagged in the slides when relevant.

Code of Conduct: All course activities, including class meetings and homework assignments, are subject to Harvard's academic integrity policies as detailed at https://handbook.fas.harvard.edu/book/academic-integrity. Please be on time and make sure that your cell phone is turned off during class time. Laptops may be used in lecture, but not in exams.

Accommodations for Students with Disabilities: If you need a disability accommodation to access this course, please communicate with me within the first two weeks of the semester. If you have your accommodation letter, please meet with me so that I can understand your needs and implement your approved accommodations. If you have not yet been approved for accommodations, please contact University Disability Services (website https://accessibility.harvard.edu; email disabilityservices@harvard.edu) to learn about their procedures.

Material Covered:

1. Descriptive statistics

   (a) Univariate descriptive statistics
   (b) Bivariate and multivariate descriptive statistics

2. Introduction to random sampling and probability
   (a) Random sampling
   (b) Variance and standard error of a sample mean
   (c) The Law of Large Numbers
   (d) Discrete and continuous distributions
   (e) Probability beyond random sampling

3. Inference on population quantities
   (a) The Central Limit Theorem
   (b) Tests, p-values, and confidence intervals

4. Causality, experimental and quasi-experimental methods
   (a) Potential outcomes and confounding
   (b) Randomized experiments
   (c) Conditional random assignment
   (d) Difference in differences
   (e) Regression discontinuity

5. Univariate regression
   (a) Derivation of Ordinary Least Squares estimator
   (b) Standard errors and confidence intervals for regression

Lecture Schedule:

- Tuesday 1/26: Lecture 1
- Thursday 1/28: Lecture 2
- Tuesday 2/2: Lecture 3
- Thursday 2/4: Lecture 4

- Tuesday 2/9: Lecture 5
- Thursday 2/11: Lecture 6
- Tuesday 2/16: Lecture 7
- Thursday 2/18: Lecture 8
- Tuesday 2/23: Lecture 9
- Thursday 2/25: Lecture 10
- Tuesday 3/2: Lecture 11
- Thursday 3/4: Lecture 12
- Tuesday 3/9: Take-home midterm
- Thursday 3/11: Lecture 14
- Tuesday 3/16: Wellness day
- Thursday 3/18: Lecture 15
- Tuesday 3/23: Lecture 16
- Thursday 3/25: Lecture 17
- Tuesday 3/30: Lecture 18
- Thursday 4/1: Lecture 19
- Tuesday 4/6: Lecture 20
- Thursday 4/8: Lecture 21
- Tuesday 4/13: Lecture 22
- Thursday 4/15: Wellness day
- Tuesday 4/20: Lecture 23
- Thursday 4/22: Lecture 24
- Tuesday 4/27: Lecture 25
- Thursday 5/6: Take-home final

Assignment and Exam Schedule:

- Tuesday 1/26: empirical exercise 1 posted
- Tuesday 2/2: problem set 1 posted
- Tuesday 2/9: empirical exercise 1 due, empirical exercise 2 posted
- Thursday 2/18: problem set 2 posted, problem set 1 due
- Thursday 2/25: empirical exercise 2 due, empirical exercise 3 posted
- Thursday 3/4: problem set 3 posted, problem set 2 due
- Thursday 3/11: take-home midterm
- Thursday 3/25: problem set 4 posted, problem set 3 due
- Thursday 4/1: empirical exercise 3 due, empirical exercise 4 posted
- Thursday 4/8: problem set 5 posted, problem set 4 due
- Thursday 4/15 : empirical exercise 4 due, empirical exercise 5 posted
- Thursday 4/22: problem set 5 due
- Thursday 4/29: empirical exercise 5 due