Statistical Learning and Causal Inference for Economics
Economics 224
Fall 2018

Course Instructor: Francis DiTraglia
Office: MCNB 535
Office Hours: M 2–4pm

Lecture Time and Location: TR 9-10:30AM, Vagelos 2000

Course Website: Course materials will be posted at `http://ditraglia.com/econ224`. You can view your grades and log onto the course discussion forum, Piazza, at `https://canvas.upenn.edu`.

Course Description: Econ 224 is an applied data analysis course that will introduce you to key ideas from modern statistical learning and causal inference, and give you hands-on experience applying them to real-world problems. Topics will include regression, randomized controlled trials, classification, instrumental variables, shrinkage methods, random forests, regression discontinuity, and differences-in-differences. Econ 224 will make heavy use of the R programming language. For more information on the topics covered, see the schedule on the final page of this document.

Active Learning: Econ 224 will not be a typical lecture-style course. Instead it will be based around "Structured, Active, In-class Learning." A typical class meeting will begin with a short quiz on the assigned readings (See "Quizzes" under "Assignments and Grading" below), followed by a mini-lecture and Q&A. The majority of class time will be devoted to working on problems and substantive data analysis "labs" in groups of 2–3 students. Grades will mainly be based on weekly problem sets and a final project. There will be no exams, but class attendance will be required. Given its somewhat unusual format, you should think carefully about whether Econ 224 is the right course for you. If you have sufficient time in your schedule to keep up with the weekly assignments, you will do well in Econ 224 and gain invaluable hands-on experience in data analysis. If you are taking a heavy course load, however, you are likely to find it difficult to keep up with Econ 224.

Prerequisites: The prerequisite for this course is Econ 103 (Statistics for Economists) or comparable coursework from the Statistics Department, e.g. STAT 430 and 431. You are expected to be conversant with elementary probability and statistics as well as the basics of R programming. For more information see "Pre-test" under "Assignments and Grading" below.

Required Texts: There are three required texts for this course:

- "Mastering 'Metrics" (MM) by Angrist & Pischke

- "An Introduction to Statistical Learning" (ISL) by James, Witten, Hastie, & Tibshirani: `http://www-bcf.usc.edu/~gareth/ISL/`

- "R for Data Science" (RDS) by Wickham & Grolemund: `http://r4ds.had.co.nz`

Note that ISL and RDS are freely available from their respective authors at the listed urls. If you prefer a physical copy, printed versions are available at the Penn bookstore and on Amazon. While MM is not available for free, it is inexpensive: $30 on Amazon at the time of this writing.

Required Software: We will use the statistical package R via a front-end called RStudio throughout the course. Both R and RStudio are free and open source. To set them up on your own computer, first download and install R from `http://cran.r-project.org/`. Then download and install RStudio by visiting `http://rstudio.org/download/desktop` and clicking the link listed under "Recommended for Your System."

Optional Texts: For students who want a deeper theoretical grounding in the material covered in Econ 224, I will assign optional readings from two additional books:

- "Mostly Harmless Econometrics" (MHE) by Angrist & Pischke

- "The Elements of Statistical Learning" (ESL) by Hastie, Tibshirani, and Friedman
  `http://www.web.stanford.edu/~hastie/ElemStatLearn/`

Note that these resources are purely optional and will not appear on problem sets or quizzes. Like its counterpart ISL, ESL is available as a free download from the authors' website.

Course Policies

Class Attendance and Participation: Because Econ 224 is an active learning course, attendance and participation are extremely important. Reading quizzes will be the primary means by which I keep track of class attendance (see "Quizzes" under "Assignments and Grading" below). As I will drop your three lowest quiz grades, you are automatically allowed three absences from class. Note that this includes absences due to illness and family emergencies. You do not have to inform me in advance, but there will be no further exceptions. Attendance does not simply mean being physically present: it means showing up prepared and participating in class activities with your group. Among other things, this means adhering to the laptop and cell phone policies listed below. You may be dismissed from class if you do not take participation seriously, as this negatively affects the others in your group. Dismissal will result in a grade of zero for that day's quiz.

Laptop Policy: You are required to bring a laptop with a working installation of R and Rstudio to each class meeting of Econ 224. If you encounter technical difficulties, please let me know. If you do not have access to a laptop, please contact me and I will make appropriate arrangements. Laptop use should be limited to course-related activities: I expect you to log out of social media and email accounts before class. You may be dismissed from class for failing to follow this policy.

Cell Phones: Cell phones are not allowed in class. I turn mine off at the beginning of each class and expect you to do the same. You may be dismissed from class for failing to follow this policy.

Email Policy: Please direct all written communication concerning Econ 224 to the course discussion forum, Piazza, rather than to the instructor or RI's personal email accounts. For questions about course material and logistics, please make your post visible to the whole class, so that others can benefit from your question and our response. (You are welcome to post anonymously.) For personal issues, use the private messaging feature to communicate directly with the instructor.

Academic Integrity: All suspected violations of the code of academic integrity as set forth in the Pennbook will be reported to the Office of Student Conduct. Confirmed violations will result in a failing grade for the course. For information about acceptable collaboration on problem sets and projects, please see the corresponding sections of "Assignments and Grading" below.

Departmental Policies: All Economics Department course policies are in force in Econ 224 even if they are not explicitly listed on this syllabus. See: `https://economics.sas.upenn.edu/undergraduate/course-information/course-policies` for full details.

Assignments and Grading

Grades for this course will be based on a pre-test, quizzes, problem sets, and a final project:

Overall Score = (10% × Pre-test) + (20% × Quizzes) + (40% × Problem Sets) + (30% × Project).

Course Curve: There will be no curve in Econ 224. This is an upper-level course that will demand a large amount of work. But provided that you put in the time and effort, you can expect to do well in the class. There should be no surprises in grading.

Pre-test: Our first class meeting of the semester, 8/28, will begin with a 20 minute pre-test covering basic statistics and R programming at the level of Econ 103. This pre-test will count for 10% of your course grade. All students who are registered for Econ 224 or are on the waiting list will recieve an email with further details on the content of the pre-test and suggestions for how to prepare. Econ 224 will be a challenging and fast-paced course and there simply isn't time for me to review material from Econ 103 during the semester. The pre-test is designed to ensure that you are adequately prepared for the course, and to help you decide if Econ 224 is right for you.

Quizzes: Every class meeting of Econ 224 other than 8/28, 11/27, 11/29, 12/04, and 12/06 will begin with a short quiz covering the assigned readings. Quizzes will count for 20% of your course grade. The purpose of these quizzes is twofold: first to give you an incentive to keep up with the readings, and second to focus your attention on the most important material. I will distribute a list of questions in advance along with each reading assignment. The quiz will simply be a random subset drawn from this list, so if you do the readings and make sure that you know how to answer each question, you will get a perfect score. You are welcome and indeed encouraged to discuss the reading questions with your fellow students and TAs either in person or on Piazza. There will be no make-up quizzes, but I will drop your three lowest scores to automatically account for absences due to illness, family emergencies, etc.

Problem Sets: With the exception of the last two weeks of the semester, I will assign a problem set at the end of class each Thursday, based on material from that week's in-class lab activities. In total there will be 11 problem sets over the course of the semester. I will calculate your grade by dropping your lowest problem set score and weighting all of the remaining ones equally. Each problem set is due at 11:59pm on Sunday night, four days after it was assigned, and must be submitted electronically via canvas. Problem set solutions should be submitted as a pdf generated from RMarkdown, and should include the accompanying RMarkdown file used to create them. (Lab 1 will explain RMarkdown and how to use it to generate pdf documents.) You may discuss problem set questions and how to solve them with your fellow students, but your code and write-up must be your own. Specifically, I expect you to adhere to an "empty hands policy." If you meet another student to discuss the problem set, you should leave the room with "empty hands," i.e. no written or digital notes of your discussion. In particular, this means that you may not share code files with one another. If you discuss the problem set with your classmates, please indicate which students you discussed with at the top of your write-up. Failing to adhere to this collaboration policy constitutes a violation of academic integrity. See "Academic Integrity" under "Course Policies."

Final Project: You and your group are required to devise and carry out a final project on a topic of your choice over the course of the semester. This project, worth 30% of your course grade, is due at 11:59pm on Sunday, December 9th. There will be no quizzes or problem sets during the last two weeks of Econ 224 to give you extra time to complete your project. Further details and requirements for final projects will be provided mid-way through the semester.

TUESDAY	THURSDAY
Aug 28th **1** Course Outline/Policies, Pre-test Lab 1: Gapminder Dataset	30th **2** Intro. to Prediction and Classification (ISL Ch. 2) Lab 2: Gapminder Dataset
Sep 4th **3** Intro. to Causal Inference I (MM 1.1) Lab 3: Racial Bias in the Labor Market	6th **4** Intro. to Causal Inference II (MM 1.2, Appendix) Lab 4: Racial Bias in the Labor Market
11th **5** Linear Regression for Prediction I (ISL 3.1–3.2) Lab 5: Predicting Election Outcomes	13th **6** Linear Regression for Prediction II (ISL 3.3–3.5) Lab 6: Crime and House Prices in Boston
18th **7** Regression and Causality I (MM 2.1–2.2) Lab 7: Class Size and Student Achievement	20th **8** Regression and Causality II (MM 2.3, Appendix) Lab 8: Class Size and Student Achievement
25th **9** Logistic Regression (ISL 4.1–4.3) Lab 9: Contaminated Drinking Water in Bangladesh	27th **10** Using Bayes' Theorem for Classification (ISL 4.4) Lab 10: Predicting Individual Ethnicity
Oct 2nd Reserve Lecture	4th **Fall Break − No Class**
9th **11** Instrumental Variables I (MM 3.1–3.2) Lab 11: Institutions and Economic Development	11th **12** Instrumental Variables II (MM 3.3) Lab 12: Institutions and Economic Development
16th **13** Cross-validation and the Bootstrap (ISL Ch. 5) Lab 13: Predicting Credit Card Default	18th **14** Model Selection (ISL 6.1) Lab 14: Baseball Player Salaries
23rd **15** Shrinkage Methods (ISL 6.2, 6.4) Lab 15: Predicting College Applications	25th **16** Tree-Based Methods (ISL 8.1 [1 & 3], 8.2 [1 & 2]) Lab 16: ???
30th **17** Regression Discontinuity I (MM 4.1) Lab 17: ???	**Nov 1st** **18** Regression Discontinuity II (MM 4.2) Lab 18: ???
6th **19** Differences-in-Differences I (MM 5.1) Lab 19: Minimum Wages and Unemployment	8th **20** Differences-in-Differences II (MM 5.2) Lab 20: Minimum Wage and Unemployment
13th **21** PCA and PCR (ISL 6.3.1, 10.1–10.2) Lab 21: Diffusion Index Forecasting	15th **22** Reserve Lecture
20th Reserve Lecture	22nd **Thanksgiving − No Class**
27th **23** Mini-lecture: Guest Presenter (TBA) Lab: Final Projects	29th **24** Mini-lecture: Guest Presenter (TBA) Lab: Final Projects
Dec 4th **25** Mini-lecture: Guest Presenter (TBA) Lab: Final Projects	6th **26** Mini-lecture: Guest Presenter (TBA) Lab: Final Projects