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 using the R programming language. For a list of topics covered in the course, see the tentative semester schedule on the last page of this document.

Active Learning: Econ 224 will be based around "Structured, Active, In-class Learning." Grades will mainly be based on weekly problem sets and a final project. There will be no exams, but class attendance will be required. The workload for Econ 224 will be fairly high, approximately 8–12 hours of time spent outside of class per week. A typical class meeting will begin with a short quiz on the assigned readings, possibly followed by a mini-lecture. The majority of class time will be devoted to working on problems and substantive data analysis "labs" in groups of 3–4 students.

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.

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`

ISL and RDS are freely available from urls listed above. Printed versions are available at the Penn bookstore and on Amazon. MM is not available for free, but costs only $30 on Amazon.

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 from `http://rstudio.org/download/desktop`.

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. 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. If you take class participation seriously, your course grade will reflect this.

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. Your participation grade may suffer if you fail 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. Your participation grade may suffer if you fail 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:

$$\text{Grade} = (20\% \times \text{Participation}) + (20\% \times \text{Quizzes}) + (30\% \times \text{Problem Sets}) + (30\% \times \text{Project}).$$

Course Curve: There will be no curve in Econ 224. This course will demand a large amount of work, but provided that you put in the time and effort, you will do well.

Participation: While we will sometimes explicitly assign participation credit for completing a non-graded task, e.g. completing an online class survey, participation credit will mainly be at the discretion of the instructor and RI. If you follow the class attendance and participation policy listed above, you will get 100% for participation. Given the active learning nature of this course, unexcused absences will count against your participation score.

Quizzes: Every class meeting with a reading assignment listed on the semester plan (see page 4) will begin with a short quiz. 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 post a list of questions on the course website 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 four lowest scores.

Problem Sets: I will assign ten problem sets over the course of the semester: in weeks 1–5 and 7–11. Each problem set is due at 11:59pm on the Sunday night after it was assigned. Problem set solutions should be submitted electronically to *canvas* and include both an `.html` file generated from RMarkdown and the `.Rmd` file used to create it. (In week 1 we will explain RMarkdown and how to use it to generate html documents.) Your grade for this component of the course will be calculated by averaging your 8 highest problem set scores 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.

Final Project: The most important part of Econ 224 is the final project. This is an opportunity for you to show what you have learned in the course by carrying out a substantive research project on a topic of your choice. Final projects will be carried out in groups of 3–4. You are welcome to form your own group; if you do not wish to form your own group, we will be happy to assign one for you. Six class meetings have been set aside for you to work on your group projects and get help from the instructor and RI: see the tentative semester plan on page 4 of this document. You will also be expected to work on the project outside of class. There will be no quizzes or problem sets in weeks 6 and 12–14 giving you ample time to work on your projects. Projects are due at 11:59pm on Sunday, December 2nd. In our final two class meetings of the semester, you and your group will give a short presentation sharing what you learned from your project. Full details and requirements for the final projects will be provided before Fall Break.

Tenatative Semester Plan

This page will be updated during the semester.

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