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, location TBA

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

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.

Course Description:

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 the basics of probability and statistics, including confidence intervals, hypothesis testing, and simple linear regression.

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 quite inexpensive: around $30 new or $20 used on Amazon. Because MM contains many equations

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. Installation instructions appear on the last page of this syllabus. RStudio is also available in the Undergraduate Data Analysis Lab (UDAL) in McNeil rooms 104 and 108–9.

Departmental Course Policies: All Economics Department course policies are in force in Econ 103 even if not explicitly listed on this syllabus. See: `http://economics.sas.upenn.edu/undergraduate-program/course-information/guidelines/policies` for full details.

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.

Piazza: We will be using an online discussion forum called Piazza, accessible via Canvas, for all written communication in this course. We will use Piazza to make course announcements, answer

questions about course material and respond to private messages from individual students regarding personal issues. By asking your question and getting an answer on Piazza, you create a positive externality: other students benefit from your questions and you benefit from theirs. You can even post anonymously if asking questions publicly makes you uncomfortable. The instructor and RIs will actively moderate Piazza both to answer questions and approve (or correct) answers written by your fellow-students. As mentioned above under "Email Policy," all written communication for Econ 103 should be directed to Piazza, not to the instructor's personal email accounts.

Homework:

Assignments and Grading

Grades for this course will be determined based on ??? Specifically,

$$\text{Overall Score} = (30\% \times \text{Quizzes}) + (20\% \times \text{Midterm 1}) + (20\% \times \text{Midterm 2}) + (30\% \times \text{Final}).$$

Course Curve:

Quizzes:

Exams:

Regrade Requests: Exam regrade requests must be made in writing within a week of receiving your graded exam. As we re-grade the entire exam, your score could rise or fall. You may not discuss your answers with an RI or the instructor before submitting a regrade request.

Installing R and RStudio

First, download and install R from `http://cran.r-project.org/`. Second, download and install RStudio by visiting `http://rstudio.org/download/desktop` and clicking the link listed under "Recommended for Your System."

Tuesday	Thursday
Aug 28th **1** Course Outline/Policies, Pre-test Lab 1: R Review, RMarkdown Overview	30th **2** Intro. to Prediction and Classification (ISL Ch. 2) Lab 2: Graduation Rates at US Colleges
Sep 4th **3** Intro. to Causal Inference I (MM 1.1) Lab 3: Data Visualization with `ggplot`	6th **4** Intro. to Causal Inference II (MM 1.2, Appendix) Lab 4: Racial Discrimination 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: Transforming Data with `dplyr`	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: Exploratory Data Analysis	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: American Housing Survey
30th **17** Regression Discontinuity I (MM 4.1) Lab 17: TBD	Nov 1st **18** Regression Discontinuity II (MM 4.2) Lab 18: Class Size and Student Achievement Redux
6th **19** Differences-in-Differences I (MM 5.1) Lab 19: TBD	8th **20** Differences-in-Differences II (MM 5.2) Lab 20: Minimum Wage and Unemployment
13th **21** Principal Components Analysis (ISL 10.1–10.2) Lab 21: US Crime Data	15th **22** Clustering Methods (ISL 10.3) Lab 22: The Fox News Effect - Media Bias and Voting
20th Reserve Lecture	22nd **Thanksgiving – No Class**
27th **23** Final Projects I	29th **24** Final Projects II
Dec 4th **25** Final Projects III	6th **26** Final Projects IV