

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-on to 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 these ideas 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.

Note that these resources are purely optional and will not appear on problem sets or quizzes.

Course Policies

Class Attendance and Participation: Because Econ 224 is an active learning course, class attendance and participation are extremely important.

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: <http://economics.sas.upenn.edu/undergraduate-program/course-information/guidelines/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:

$$\text{Overall Score} = (10\% \times \text{Pre-Test}) + (20\% \times \text{Quizzes}) + (40\% \times \text{Problem Sets}) + (30\% \times \text{Project}).$$

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 receive 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 help you decide if Econ 224 is the right course for you before the end of the add/drop period.

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 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:

Final 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.

TUESDAY	THURSDAY
<div>Aug 28th</div> <div>1</div> Course Outline/Policies, Pre-test Lab 1: Dynamic Documents with RMarkdown	<div>30th</div> <div>2</div> Intro. to Prediction and Classification (ISL Ch. 2) Lab 2: Graduation Rates at US Colleges
<div>Sep 4th</div> <div>3</div> Intro. to Causal Inference I (MM 1.1) Lab 3: Data Visualization with <code>ggplot</code>	<div>6th</div> <div>4</div> Intro. to Causal Inference II (MM 1.2, Appendix) Lab 4: Racial Discrimination in the Labor Market
<div>11th</div> <div>5</div> Linear Regression for Prediction I (ISL 3.1–3.2) Lab 5: Predicting Election Outcomes	<div>13th</div> <div>6</div> Linear Regression for Prediction II (ISL 3.3–3.5) Lab 6: Crime and House Prices in Boston
<div>18th</div> <div>7</div> Regression and Causality I (MM 2.1–2.2) Lab 7: Transforming Data with <code>dplyr</code>	<div>20th</div> <div>8</div> Regression and Causality II (MM 2.3, Appendix) Lab 8: Class Size and Student Achievement
<div>25th</div> <div>9</div> Logistic Regression (ISL 4.1–4.3) Lab 9: Contaminated Drinking Water in Bangladesh	<div>27th</div> <div>10</div> Using Bayes' Theorem for Classification (ISL 4.4) Lab 10: Predicting Individual Ethnicity
<div>Oct 2nd</div> <div></div> Reserve Lecture	<div>4th</div> <div></div> Fall Break – No Class
<div>9th</div> <div>11</div> Instrumental Variables I (MM 3.1–3.2) Lab 11: Exploratory Data Analysis	<div>11th</div> <div>12</div> Instrumental Variables II (MM 3.3) Lab 12: Institutions and Economic Development
<div>16th</div> <div>13</div> Cross-validation and the Bootstrap (ISL Ch. 5) Lab 13: Predicting Credit Card Default	<div>18th</div> <div>14</div> Model Selection (ISL 6.1) Lab 14: Baseball Player Salaries
<div>23rd</div> <div>15</div> Shrinkage Methods (ISL 6.2, 6.4) Lab 15: Predicting College Applications	<div>25th</div> <div>16</div> Tree-Based Methods (ISL 8.1 [1 & 3], 8.2 [1 & 2]) Lab 16: American Housing Survey
<div>30th</div> <div>17</div> Regression Discontinuity I (MM 4.1) Lab 17: Tidying Messy Data	<div>Nov 1st</div> <div>18</div> Regression Discontinuity II (MM 4.2) Lab 18: Class Size and Student Achievement Redux

TUESDAY	THURSDAY
6th Differences-in-Differences I (MM 5.1) Lab 19: Scraping Data from the Web	8th Differences-in-Differences II (MM 5.2) Lab 20: Minimum Wage and Unemployment
13th Principal Components Analysis (ISL 10.1–10.2) Lab 21: US Crime Data	15th Clustering Methods (ISL 10.3) Lab 22: The Fox News Effect - Media Bias and Voting
20th Reserve Lecture	22nd Thanksgiving – No Class
27th Mini-lecture: “Causal Forests” (Davis & Heller, 2017) Lab: Final Projects I	29th Mini-lecture: “The Costs of Algorithmic Fairness” Lab: Final Projects II
Dec 4th Mini-lecture: Guest Presenter (TBA) Lab: Final Projects III	6th Mini-lecture: Guest Presenter (TBA) Lab: Final Projects IV