Course Overview

This note summarizes the objectives of Prescriptive Models and Data Analytics, provides administrative information and describes course requirements, and provides a detailed syllabus.

Course Goals

As companies invest in the accumulation of large data assets as well as experimentation platforms, a key priority is to develop the capabilities for unlocking the value of these data resources. Central to this challenge is developing the human resources to perform data analytics. The goal of this course is to enhance your ability to perform transparent, rigorous, and convincing data analytics. The specific aims are to:

(1) Learn data analytics methods with a focus on "big data" methods and experimentation and causal inference through A/B testing

(2) Gain an understanding of which strategy is best suited for solving a specific business challenge

In order to achieve those goals, the course will have a strong focus on applications. We will cover the basic statistical foundations of each method with a focus on an intuitive understanding of the relevant econometrics. In practice, we will work on a set of in-class exercises that allow you to implement each method based on real data using the statistical software R.

Topics

We will cover topics in two broad domains:

Causal Inference. In order to optimize business strategy along various dimensions such as price setting and optimal advertising budgets, we need to understand how the various variables CAUSALLY affect demand and profits. We will cover A/B tests that are increasingly employed in many industries as well as various other methods of causal inference when A/B tests are not feasible such as difference-in-differences regressions and panel data methods.

Predictive Models. Due to increasing amounts of data on consumers, business decisions often require us to find useful information in large data sets. Recent advances in machine learning allow us to handle "big data" on consumer behavior in an efficient way. We cover the problem of overfitting and how to overcome it through cross-validation, LASSO regressions, and regression trees / random forests.

Class Structure

In order to maximize your learning experience, we will follow a very specific structure in terms of pre-class preparation, work in class, and problem sets.

Preparation Before Class

There will be no required readings because classes will be self-contained. However, for most classes I will indicate on the syllabus which concepts will be particularly important in each class. Furthermore, classes will build on each other, so you should always make sure that you are familiar with all the concepts covered in past classes. You should aim to spend 30 - 60 minutes reviewing relevant concepts and past class notes before each class.

Class Structure

All classes will be split into two halves. The first half will contain lecture-based material that introduces new statistical concepts. In most lectures we will almost immediately use an actual data-set to illustrate and understand those concepts together. In the second half of the course you will work on a data task in small groups. These "workshops" will allow you to see the concepts introduced in the first half in action and apply them to a real-world business problem.

Problem Sets

There are 5 problem sets throughout the quarter (roughly every two weeks, see the last page of this document for details on deadlines). If you miss the submission deadline, you will receive a grade of zero. There will be no exception to this policy.

Optional Readings

I will provide a relatively generous list of optional readings. These are strictly optional, and exam material will not be taken directly from those readings. The readings will allow you to delve deeper into topics you find particularly interesting and can serve as a repository later if you want to learn about a specific technique when it becomes relevant for your job.

Grade Breakdown

5 Problem Sets	50 %
Final Exam	50 %

Final Exam

The final exam will be a 3 hour open book and open notes take-home exam that will be administered online.

Office Hours

tba

Optional Textbooks

There are no required textbooks for this course, but good complimentary readings are the following:

1. *Business Data Science*, By Matt Taddy (McGraw-Hill Education, 2019).
2. *Mastering 'Metrics: The Path from Cause to Effect*, by Joshua D. Angrist and Jörn-Steffen Pischke (Princeton University Press, 2015).

3. *Mostly Harmless Econometrics: An Empiricist's Companion*, by Joshua D. Angrist and Jörn-Steffen Pischke (Princeton University Press, 2009).

4. *Introduction to Econometrics (1st Edition)*, by James H. Stock and Mark W. Watson (Addison Wesley, 2003).

Among the four books, (1) is closest in terms of content to the lecture material, however the material is covered with a higher level of technical detail. (2) is a good and very easy to read book that covers the causal inference material of the course (but not the predictive modeling part). The same is true for (3), which covers the identical material as (2) but is more technical. This book is useful if you want to delve deeper into specific topics and/or would like to understand the derivation behind specific concepts. (4) is a more "standard" econometrics textbook and is less focused on applications than the other three (and also does not cover predictive modeling). It is slightly more technical and focuses more on the relevant math rather than verbal explanations.

Syllabus

---

1. January 6: ***Introduction and Uni-variate Regression***

Prepare for Class:

- Install R and RStudio and make sure they are running
- No other preparation required for the first session :)

Introduction:

- Course introduction
- Examples of impactful data analytics
- Course structure and logistics

Uni-variate regression:

- Interpreting intercept & slope coefficient
- Precision of regression estimates

Optional Reading:

- *Personalized Pricing and Customer Welfare*; J.-P. Dube, and S. Misra, Working Paper, May 2019.
- *Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment*; T. Blake, C. Nosko, and S. Tadelis, Econometrica, 83(1), February 2015.
- *Big Data and Marketing Analytics in Gaming: Combining Empirical Models and Field Experimentation*; H. S. Nair, S. Misra, W. J. Hornbuckle IV, R. Mishra, and A. Acharya, Marketing Science, 36(5).

2. January 13: ***Causality, Multi-variate regression, A/B Testing***

Causality & Multi-variate regression:

- Correlation and causality
- Omitted variable bias

A/B tests:

- A/B tests and causality
- The importance of randomization

Optional Reading:

- *Online Ads and Offline Sales: Measuring the Effect of Retail Advertising via a Controlled Experiment on Yahoo!*; R. A. Lewis, and D. H. Reiley, Quantitative Marketing and Economics, 12(3), 2014.
- *Measuring Consumer Sensitivity to Audio Advertising: A Field Experiment on Pandora Internet Radio*; J. Huang, D. H. Reiley, and N. M. Riabov, Working Paper, April 2018.

Additional Reading on Methodology:

- Mastering 'Metrics: chapter 1
- Mostly Harmless Econometrics: chapter 2

3. **January 24 (Friday, b/c of MLK holiday):** ***A/B Testing, Precision, & Sample Size***

Review the following concepts before class:

- Causality and omitted variable bias
- Standard error formula for uni-variate regression coefficient
- Regression residuals

Lecture:

- Precision of A/B tests
- Sample size and other A/B test design choices

Workshop:

- Measuring ad effectiveness
- Understand ways to influence precision

Optional Reading:

- *The Unfavorable Economics of Measuring the Returns to Advertising*; R. A. Lewis, and J. M. Rao, The Quarterly Journal of Economics, 130(4), November 2015.
- *Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment*; Dean Karlan and John A. List, American Economic Review, 97(5).

Additional Reading on Methodology:

- Mastering 'Metrics: chapter 1
- Mostly Harmless Econometrics: chapter 2

4. January 27: ***Control Variables***

Review the following concepts before class:

- Causality and omitted variable bias
- Regression residuals

Lecture:

- Isolating "good" (i.e. random) variation
- Understand how to deal with partial randomization via control variables
- Sequential implementation of multi-variate regression
- Regression mechanics: Dummy variables

Workshop:

- Estimating demand for rideshare services
- Separating demand and supply side factors
- Regression discontinuity design

Optional Reading:

- *Retention Futility: Targeting High-risk Customers Might be Ineffective*; E. Ascarza, Journal of Marketing Research, 55(1), February 2018.

Additional Reading on Methodology:

- Mastering 'Metrics: chapter 2, in particular pages 68-74
- Mostly Harmless Econometrics: chapter 3, in particular pages 59-64
- Stock & Watson: chapter 6.3

5. February 3: ***Panel Data Methods***

Review the following concepts before class:

- Material on control variables (see previous class)
- In particular: sequential estimation of multi-variate regression

---

Lecture:

- Regression mechanics: Interaction terms
- Cross-sectional and time fixed effects
- Relationship of panel methods to control variables

Workshop:

- Revolving door lobbyists: quantifying the role of political connections
- Understand the role of cross-sectional and time fixed effects

---

Optional Reading:

- *Revolving Door Lobbyists*; Jordi Blanes i Vidal, Mirko Draca, and Christian Fons-Rosen, American Economic Review, December 2012.

Additional Reading on Methodology:

- Mastering 'Metrics: chapter 5 & pages 191-196
- Mostly Harmless Econometrics: chapter 5.2
- Stock & Watson: chapter 8

6. February 10: ***Difference-in-differences Regression***

Lecture:

- General logic of difference-in-differences
- Difference-in-differences in a regression framework
- Relationship to panel data methods

Workshop:

- Measuring the impact of the Philadelphia soda tax
- Implement a difference-in-differences regression specification

---

Optional Reading:

- *The Impact of Soda Taxes: Pass-through, Tax Avoidance, and Nutritional Effects*; S. Seiler, A. Tuchman, and S. Yao, Working Paper, May 2019.
- *Does Online Word-of-Mouth Increase Demand? (and How?) Evidence from a Natural Experiment*; S. Seiler, S. Yao, and W. Wang, Marketing Science, 36(6), December 2017.

Additional Reading on Methodology:

- Mastering 'Metrics: chapter 5 & pages 191-196
- Mostly Harmless Econometrics: chapter 5.2
- Stock & Watson: chapter 8

7. **February 21 (Friday, b/c of Presidents' Day):** ***Introduction to Predictive Models***

Review the following concepts before class:

- Regression fit measures: r-squared
- (Exact!) interpretation of the p-value (of a specific regression coefficient)

Lecture:

- The role of predictive models
- The problem of overfitting
- Out-of-sample fit and cross-validation

Workshop:

- Optimally targeted search advertising
- Embed empirical analysis directly into managerial decisions
- Understand the importance of cross-validation
- Value of high-dimensional targeting for profits

Additional Reading on Methodology:

- Business Data Science: pages 69-74

8. February 24: ***Regularization & Lasso Regression***

Review the following concepts before class:

- Material from last class (all of it is essential for this class)

Lecture:

- Model selection in high-dimensional settings
- Simple methods: forward / backward iterative selection
- Regularization: Lasso, Ridge, and other penalty functions
- Implementation of Lasso regression (on search advertising data from last week)

Workshop:

- Using Lasso regression to select control variables
- Use Lasso as a scalable method for causal inference

Optional Reading:

- *Heterogeneous Treatment Effects and Optimal Targeting Policy Evaluation*; G. J. Hitsch, and S. Misra, Working Paper, February 2018.

Additional Reading on Methodology:

- Business Data Science: Chapter 3

9. **March 2:** ***Causal Lasso & Regression Trees and Forests***

Lecture:

- Causal lasso
- Non-parametric variable selection via regression trees
- Random forests

Workshop

- Algorithmic advertising targeting

Additional Reading on Methodology:

- Business Data Science: Chapter 9

10. **February 9:** ***Re-cap & Exam Prep***

Which method to use for causal inference:

- A/B tests
- Panel data methods / Difference-in-differences
- Control variables (possibly in combination with Lasso)

Predictive methods

- Targeting & treatment effect heterogeneity
- Predictive questions (no manipulation involved)
- Improve precision in A/B tests

When to use purely predictive models versus causal models

- Which marketing questions require which type of approach?

Final Exam Review

Problem Set Timeline

Problem Sets are due Friday at 11:59pm (except for the final problem set) and have to be submitted via CCLE.

Friday January 17th: Problem Set 1: A/B Tests

Friday January 31st: Problem Set 2: Control Variables

Friday February 14th: Problem Set 3: Panel Data & Diff-in-diff

Friday February 28th: Problem Set 4: Lasso Regression

Monday March 9th, 4pm: Problem Set 5: Exam Prep Problem Set