Draft Syllabus
University of Pennsylvania
The Wharton School
Fall 2018

BEPP 280: APPLIED DATA ANALYSIS

Classroom: TBD   **Time**: Monday/Wednesday, 3 pm to 4:20 pm

Instructor: Santosh Anagol   **Email**: anagol@wharton.upenn.edu
Office: SH-DH 1458   **Phone**: (215) 898-8920
Office hours: TBD
Teaching Assistant: TBD

<u>**Description**</u>: The most important questions, debates and decisions in business and public policy revolve around understanding whether some variable X affects some other variable Y. For example:

- Does the minimum wage affect employment?
- Do taxes affect economic growth?
- How does demand for a product respond to changes in price?
- Does advertising increase demand?

The same challenges arise in attempting to evaluate the impact of a program or policy on an outcome of interest. However, as is well known, a simple correlation between X and Y does not mean that X causes Y, or that if you as a businessperson or policy-maker change X, then Y will change. It may instead be that Y affects X; some third factor Z affects both X and Y; or there could be a completely artificial or spurious correlation or trend between X and Y, with neither affecting the other at all.

This course will examine how and when data can be used specifically to infer whether there is a causal relationship between two variables. We will emphasize (a) the critical role of an underlying economic theory of behavior in interpreting data and guiding analysis, as well as (b) a range of advanced techniques for inferring causality from non-experimental data, such as randomized controlled trials, regression discontinuity, difference-in-difference, audit (mystery shopping) approaches and stock-market event studies.

The issue of causality, and the relevance of thinking about models and methods for inferring causality, is just as central and important for "Big Data" as it is when working with traditional data sets in business and public policy. The emphasis will not be on proofs and derivations but rather on understanding the underlying concepts, the practical use, implications and limitations of techniques. Students will work intensively with data, drawing from examples in business and public policy, to develop the skills to use data analysis to make better decisions. All analysis will be conducted using R. The goals of the course are for students to become expert consumers able to interpret and evaluate empirical studies as well as expert producers of convincing empirical analysis themselves.

<u>**Course Materials:**</u> We will use the following book throughout the course: Angrist, Joshua and Jörn-Steffen Pischke (2014). <u>Mastering 'Metrics</u>. Princeton University Press.

Readings will also include selected journal articles. Finally, students who are not already completely fluent in the basic linear regression model from previous statistics courses should consult a textbook along with the course. Some useful and readable ones are: Wooldridge, Jeffrey, Introductory Economics: A Modern Approach, South-Western College Publishers, any edition (you should be comfortable with the material in chapters 2-5 at a minimum), and James Stock and Mark Watson, Introduction to Econometrics (3rd ed.), Addison Wesley (chapters 4-7).

Pre-requisites: Relevant introductory statistics course covering at least hypothesis testing and linear regression (STAT 102, 112 or 431 or equivalent), or at the discretion of the instructor. The TA will give a couple crash courses in R at the beginning of the semester, to help students get up to speed for using R to complete the problem sets.

Grades and Assignments: There will be five problem sets, two in-class exams and a group presentation. The schedule below provides the corresponding dates. No credit will be given for late problem sets. The weighting of these assignments in the final grade will be as follows:

Assignment	% of Grade	Due Date
First Exam (Sessions 1-14)	40%	
1st Problem Set	5%	
2nd Problem Set	5%	
3rd Problem Set	5%	
4th Problem Set	5%	
Second Exam (Sessions 15-27)	40%	

Exams: Both exams will be closed book. The first exam will focus on the lectures 1-14 and the second will focus on the entire semester, though with more of a focus on the material from lectures 16-27.

Students will be given a period of one week from the date the exam is returned to request a regrade. The BEPP Department maintains a photocopy of all exams. All requests must be made in writing, stating the reason they believe they should receive a regrade no later than the close of business on the day which is one week after the exam is made available. Note that the entire exam will be reviewed during a regrade, not only the questions/answers requested, and a regrade may be higher or lower than the original grade.

Problem Sets: Students may work in groups of 2 or 3, but submit their own version. Problem sets will involve analyzing data using the techniques learned in class. Problem sets will be submitted and graded via Canvas.

Code of Academic Integrity: All students enrolled in courses in the Business Economics and Public Policy Department are expected to comply with the University of Pennsylvania's Code of Academic Integrity. We encourage all students to read the Code so that they are well aware of all situations that would be considered a violation. It is the policy of the Department of Business Economics and Public Policy to immediately fail any student who is to be in violation of the Code. Cheating, in any manner, on a graded assignment or exam will result in failing both the assignment/exam and the course. In addition to the sanctions imposed by the Department of Business Economics and Public Policy, the Office of Student Conduct may impose additional sanctions. Please review the Code of Academic Integrity as well as example of violations and possible sanctions: http://www.upenn.edu/provost/PennBook/academic_integrity_code_of

Support Functions: Beth Moskat (emoskat@wharton.upenn.edu) will provide classroom support.

Lecture 1. The Value of Experiments

- Angrist and Pischke, Introduction
- Run Field Experiments to Make Sense of Your Big Data, Harvard Business Review, 11/12/15
- How to Design Smart Business Experiments, Harvard Business Review, 2/2009
- The Discipline of Business Experimentation, Harvard Business Review, 12/2014

Lecture 2. The Randomized Control Trial (RCT) Design (aka the A/B Test Design)

- Angrist and Pischke, Chapter 1

Problem Set 1 Due on Canvas

Lecture 4. RCT Application 1: Virgin Atlantic Experiments to Help the Environment

- Virgin Atlantic Tested 3 Ways to Change Employee Behavior, Harvard Business Review,
- List, Et. al., A New Approach to an Age Old Problem: Solving Externalities by Incenting Workers Directly

Lecture 5. RCT Application 2: Does Working from Home Work?

- A Working From Home Experiment Shows High Performers Like it Better
- To Raise Productivity, Let More Employees Work from Home

Lecture 6. RCT Application 3: Testing for the Endowment Effect Among Investors

- To Have and To Hold, The Economist, 8/18/16
- Voxeu, "Endowment Effects in the Field: Evidence from IPO Lotteries in India", 7/7/16

Lecture 7. RCT Application 4: Should NGOs Charge Money for Life-Saving Technologies?

- Dupas, P. and J. Cohen "Free-Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment"

Lecture 7: The Regression Design (Review)

- Angrist and Pischke, Chapter 2
- Dale, and Krueger. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables"

Lecture 8: What to Do When the Treatment Group Isn't Treated? RCT and Instrumental Variables

- Angrist and Pischke, Chapter 3

Lecture 9: Application of Instrumental Variables in RCT

- Location, Location, Location: Repetition and Proximity Increase Advertising Effectiveness
- People and Cookies: Imperfect Treatment Assignment in Online Experiments

Lecture 10. The Regression Discontinuity Design

- Angrist and Pischke, Chapter 4

Lecture 11. Regression Discontinuity Application 1: the Value of Uber to Consumers

- Using Big Data to Estimate Consumer Surplus: The Case of Uber

Problem Set #2 Due

Lecture 12. Regression Discontinuity Application 2: The Effect of Consumer Reviews on Business

- HBS Working Knowledge: The Yelp Factor: Are Consumer Reviews Good for Business?
- Reviews, Reputation, and Revenue: The Case of Yelp.com

Lecture 13. Regression Discontinuity Application 3: Rank Effects in Elections and More

- Lee, Randomized Experiments from Non-Random Selection in US House Elections
- Anagol and Fujiwara, The Runner Up Effect

Lecture 14: Synthesis and Review for First Exam

Lecture 15. In-Class Exam

Lecture 16. The Difference-in-Difference Design

- Angrist and Pischke, Chapter 5

Lecture 17. Diff-in-Diff Application 1: Does the Minimum Wage Reduce Employment?

Lecture 18. Diff-in-Diff Application 2: Advertising Experiments at the Ohio Art Company

- Advertising Experiments at the Ohio Art Company

Lecture 19. Diff-in-Diff Application 3: Did Beer Get More Expensive When Coors/Miller Joined Forces?

- Ashenfelter, et. al. "Efficiences Brewed: Pricing and Consolidation in the US Beer Industry"

Lecture 20. The Audit Study Design (aka Mystery Shopping)

Problem Set #3 Due

Lecture 21. Audit Studies of Discrimination

- Bertrand and Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination."

Lection 22. The Stock Market Event Study Design

- Intro to Stock Market Event studies
- Dividends

Lecture 23. Stock Market Event Study Application 1: Was that Merger A Good Idea?

- Winning by Losing: Evidence on Overbidding in Mergers (with E. Moretti and F. Peters), August 2016. Revise and resubmit at the *Review of Financial Studies*

Lecture 24. Stock Market Event Study Application 2: How Politics Affects Stock Prices and Corporate Value

- Fisman, R. "The Value of Political Connections"

- Fisman, R. "The Value of Connections to Vice President Cheney"
- Fisman, R. and Zitzewitz, E. "Value of Trump Presidency"
- Acemoglu et. al. "The Value of Connections to Tim Geithner"

Lecture 25. Machine Learning and Prediction Problems in Business and Policy

- Mullainathan Prediction Problems and Motivation
- Lasso Intro

Problem Set #4 Due

Lecture 26. Applied Prediction Problem: To Jail or Not to Jail?

- Mullainathan jail paper
- Poverty prediction

Lecture 27. Synthesis and Review for Second Exam

Lecture 28. In-Class Exam