University of Pennsylvania
The Wharton School
Fall 2022

BEPP 2800: APPLIED DATA ANALYSIS AND CAUSALITY FOR BUSINESS AND PUBLIC POLICY

Instructor: Santosh Anagol **Email**: anagol@wharton.upenn.edu
Classroom:
Section 001: TR 1:45 pm – 3:15 pm
Anagol Office hours: Friday 2:00 pm to 3:00 pm

Teaching Assistant:

Teaching Assistant Office Hours:

Description: 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:

- How do website design features change demand for products and government services?
- Does allowing workers to work from home increase or reduce productivity?
- How does demand for a product respond to changes in price?

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) how experiments (A/B testing etc.) can be used to answer questions in business and public policy (b) a range of advanced techniques for inferring causality from non-experimental data, such as regression discontinuity, regression-kink, difference-in-difference, synthetic control, event-study, and audit (mystery shopping) approaches.

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 the freely available R language, which is widely used in business and policy. 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. To that end, this is a hands-on course with the main deliverables being problem sets and a group project where you will apply the techniques in class to data.

Please note that while STAT 490 “Causal Inference” has broadly similar motivations, the courses were designed to cover substantially different techniques. Students interested in understanding causal relationships are encouraged to take both.

This course counts towards the Business Analytics and BEPP concentrations, as well as fulfilling the Technology, Innovation and Analytics requirement.

Pre-requisites: Relevant introductory statistics course covering at least hypothesis testing and linear regression (STAT 102, 112 or 431 or equivalent (i.e. Econ 104, PSCI 338, SOC120)), or at the discretion of the instructor. The TA will give a crash course in R at the beginning of the semester, to help students get up to speed for using R to complete the problem sets and in-class demos.

Course Materials: The primary source for this course are the in class lectures. The required and optional readings will be placed in Canvas folders under the Readings folder. There is no required textbook, but for those who would like to have a textbook on hand, excellent books covering much of the course material are: Gerber and Green's "Field Experiments: Design, Analysis, and Interpretation" and Angrist and Pischke's "Mastering 'Metrics." These can be relied on for questions about classwork, homework, and your natural experiment project. These books are on reserve at Lippincott library.

For brushing up on basic regression, Chapter 2 of Mastering Metrics should be your first choice. It may also be useful to consult: Wooldridge, Jeffrey, Introductory Econometrics: 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).

Grades and Assignments: The schedule below provides the corresponding dates for deliverables. The weighting of these assignments in the final grade will be as follows:

Assignment	% of Grade	Due Date
Class participation	5%	
1st Graded Problem Set	5 %	
2nd Graded Problem Set	5 %	
First Exam	35%	
Natural Experiment Project	15%	
Second Exam	35%	

Class Participation: I expect students to be prepared for class and actively participate.

Problem Sets: I will assign three problem sets. The highest two scores will count towards the students' grade (i.e. there is no penalty for not submitting a third problem set). Students may work in groups of 2 or 3, but all students must submit their own solutions. Problem sets will involve analyzing data using the techniques learned in class. Problem sets will be submitted and graded via Canvas. Submissions are scored and returned via Canvas. Partial credit is given for problem sets turned in within 8 hours of the normal deadline in the following manner (determined by time-stamp for upload to Canvas): 20% reduction for *any* lateness, plus a reduction of 10% per hour, giving partial credit to any assignment turned in before 8:00AM the next day. (E.g., the score for an assignment turned in at 2:30AM would be reduced by 20%+25% = 45%.) Students are encouraged to turn in problem sets prior to the due date.

Natural Experiment Project: Students will self-organize in to groups of 4 -5 students (or 3 if necessary based on final enrollments) and create a 20 minute video presentation of a real data analysis of a natural experiment. Groups must be formed within your section of BEPP 280. There will be a section specific discussion board where students can list interests regarding topics to help facilitate group formation, and Prof. Anagol will assign you a group if helpful.

Each group will prepare a very short, informal proposal for a topic/project idea, and will work together with classmates to develop this during two class sessions in breakout groups (November 23 and 30). During the week of December 2 and December 9 there will be no class, instead there will be sign-ups for 20 minute slots to discuss your project progress with Prof. Anagol. It is mandatory that each group meet with Prof. Anagol at least once, and two meetings are encouraged. Please see the canvas document "BEPP 280 Natural Experiment Instructions" for further details.

After all groups have submitted each group has submitted their videos, each student will choose five other groups' videos to watch and submit a feedback form on Canvas. These feedback forms will be due at 5 pm on December 15. Each feedback form completed will be worth .5 points toward you total 5 point class participation grade.

R Resources at Penn:

Sarina Divan (the TA for this course) will hold weekly office hours to help with R and problem sets. You should also feel free to email her with R related questions.

The Weigle Information Commons' Student Assistance Services offers appointments for Statistical Software Consultations to receive assistance and guidance on R and other stat programs. Hours are in WIC Room 128 by appointment using this online scheduler http://libcal.library.upenn.edu/appointment/36597.

More information on R resources at Penn: http://guides.library.upenn.edu/stat_packages/R

Exams: The exams will be closed book and cover the main themes discussed throughout the class. The second exam will focus mainly on the material from the second half of the course.

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.

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	Date	3 Topic
1	30-Aug	Why Causality Matters
2	1-Sep	Randomized Control Trials (A/B Testing)
3	6-Sep	Experimental Analysis Using Regression
4	8-Sep	Precision, Power and Sample Size (**R Demo**)
5	13-Sep	Randomization, Blocking, and Balance (**R Demo**)
6	15-Sep	Non-Compliance (**R Demo**)
Deliverable	*20-Sep*	*PS 1: RCT and quantifying uncertainty problem set due at midnight on Canvas.*
7	20-Sep	Placebo and Hawthorne Effects
8	22-Sep	Sample Attrition (**R Demo**)
9	27-Sep	Heterogeneous Treatment Effects (**R Demo**)
10	29-Sep	First Exam Review
11 & Deliverable	*4-Oct*	First Exam
12	11-Oct	The Regression Discontinuity Design (**R Demo**)
13	13-Oct	RD Application 1: Demand Curve Estimation (**R Demo**)
14	18-Oct	RD Application 2: Border Discontinuity Design
Deliverable	20-Oct	*PS 2: Regression Discontinuity Design due at midnight on Canvas*
15	20-Oct	RD Application 3: Attention and Heuristic Effects
16	25-Oct	The Difference in Difference Design (DiD)
17	27-Oct	The Difference in Difference Design with Multiple Treatments (**R Demo**)
18	1-Nov	Stock Market Event Study Design (**R Demo**)
19	3-Nov	Class Cancelled
20	8-Nov	The Synthetic Control Design
Deliverable	10-Nov	*PS 3: Difference-in-Difference Problem Set Due at midnight on Canvas*
21	11-Nov	Synthetic Control Design in R (**R Demo**)
22	15-Nov	Synthesis and Review for Second Exam
23 & Deliverable	17-Nov	Second Exam
24	22-Nov	In-Class Group Breakout Rooms to Determine Feasible Natural Experiment Project
25	29-Nov	In-Class Breakout Rooms Plan for Gathering Data & Analysis
26	1-Dec	No class: Sign-ups this week 20 min feedback sessions on project from Prof Anagol
27	6-Dec	No class: Sign-ups this week 20 min feedback sessions on project from Prof Anagol
28	8-Dec	No class: Sign-ups this week 20 min feedback sessions on project from Prof Anagol
Deliverable	13-Dec	Natural Experiment Video Presentation Due at 5 pm
Deliverable	16-Dec	Presentation Feedback on Chosen 5 projects due on Canvas at 5 pm