Overview and Class Goals
This is the second course in a three-course sequence on quantitative political methodology, by which we mean the application of statistical methods to problems in political science and public policy. The goal of the three-course sequence is to teach you (1) to understand and (2) to confidently apply a variety of statistical methods and research designs that are essential for political science and public policy research.

Building on the first course (17.800) which covered regression models, this second class provides a survey of more advanced empirical tools for political science and public policy research. The focus is on statistical methods for causal inference, i.e. methods designed to address research questions that concern the impact of some potential cause (e.g., an intervention, a change in institutions, economic conditions, or policies) on some outcome (e.g., vote choice, income, election results, levels of violence).

We cover a variety of causal inference designs, including experiments, matching, regression, panel methods, difference-in-differences, synthetic control methods, instrumental variable estimation, regression discontinuity designs, quantile regressions, and bounds. We will analyze the strengths and weaknesses of these methods. Applications are drawn from various fields including political science, public policy, economics, and sociology.

The class is open to qualified students from other departments and undergraduates.

Prerequisites
A willingness to work hard on possibly unfamiliar material. In addition to introductory statistics and probability, the course assumes a good knowledge of linear regression meaning that you should have taken at least one graduate class on this subject (such as 17.800). Students are also expected to be reasonably proficient in the statistical software R (you may use other software packages that you are very familiar with, but we can only support R).

Class Requirements
Reading
The syllabus lists the required readings for every week. This required reading should be completed prior to lecture in a given week. Students are expected to read the material very carefully. You may even find it helpful to read the material multiple times.
Homework
This is a methodological course, developing skills in understanding and applying statistical methods. You can only learn statistics by doing statistics and therefore the homework for this course is extensive, including weekly homework assignments. The assignments consist of analytical problems, computer simulations, and data analysis. They will usually be assigned on Tuesday night and due the following Tuesday, prior to lecture. No late homework will be accepted. All sufficiently attempted homework (i.e. a typed and well organized write-up with all problems attempted) will be graded on a (+,✓,-) scale. We encourage students to work together on the assignments, but you always need to write your own solutions, and we ask that you make a solo effort at all the problems before consulting others. We also ask that you write the names of your co-workers on your assignments.

Student Project
Students are expected to write a short empirical paper that applies methods learned in this class to a research question of their choice. The paper should be 5-15 pages in length and focus on the research question, data, empirical strategy, results, and conclusions. Literature reviews, background, lengthy motivations, etc. should be omitted or may be included as an appendix. You also need to submit a copy of your analysis code. Students are free to choose any topic they want, as long as they have a clear research question that concerns the causal effect of some institution, intervention, policy, or event on some outcome, result, or performance. Co-authored projects are very strongly encouraged (learning to co-author is essential because nowadays most articles in political science are co-authored).

Students need to meet the following milestones for their project:

• By 3/15: Email the instructor and TA a short description of your proposed project (i.e. half a page). Students are encouraged to meet with the instructor and TA during office hours to discuss their project and progress.

• By 4/12: Email the instructor and TA a 2-3 page description of progress, analysis, and preliminary results.

• By 5/8: Email the instructor, TA, and the entire class the first draft of your project. Everyone is expected to read all these submissions prior to the student presentations that follow.

• On 5/10 and 5/15: Students will present their projects to the class. Presentations will be no more than 5-10 minutes in length, and will be oral with the aid of 2-5 slides that summarize the main results.

• By 5/17: Email the instructor and TA the final version of your project.

Grading
Grades will be based on

• weekly homework assignments (65% of final grade)

• student project (30% of final grade)

• participation and presentation (5% of final grade).

Recitation Sections
Weekly recitation sections will be held on Friday. The section will cover a review of the theoretical material and also provide help with computing issues. The TA will run the sections and can give more detail.

Computation
We teach the course in R, which is an open-source computing language that is very widely used in statistics. You can download it for free from www.r-project.org. The web provides many great tutorials and resources to learn R. A list of these is provided at http://wiki.math.yorku.ca/index.php/R:Getting_started. A nice way to start you off are the two video tutorials provided by Dan Goldstein here and also here. R runs on a wide variety of UNIX platforms, Windows and MacOS. R makes programming very easy, has strong graphical
capabilities, and also contains canned functions for most commonly used estimators.

To refresh your R you are expected to work through one of the following free tutorials unless you are well familiar with this material. All three tutorials cover similar material, just pick the one you like best:


If you are very familiar with another statistical software package you may use that for the course at your own risk. We can only support R.

**Course Website**

**Course Forums**
The course website has a discussion board in the “forum” section. This discussion board provides an opportunity to post questions regarding the course material and/or computing. In addition to precepts and office hours, please use this Forum on the Stellar course website when asking questions about lectures, problem sets, and other course materials. This will allow students to see other students’ questions and learn from them. Both the TA and the instructor will regularly check the Board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. You can also sign up to receive notifications for posted questions and answers. A student’s respectful and constructive participation on the forum will count toward his/her class participation grade. Do not email your statistical questions directly to the instructor (unless they are of personal nature) — we will not answer them!

**Schedule**
Please notice the following scheduling issues:

- No class on 2/21 (Due to Presidents Day on Monday, Tues 2/21 will have a Monday schedule),
- No class on 3/27 and 3/29 (Spring Break).
- No class on 4/17 (Patriots Day).

**Books**

**Main Books**

- We will read chapters from the following textbooks which are available at the COOP and also on reserve in the library.

**Useful Summary Articles**

- The following papers summarize the main methods learned in this course. They are dense and detailed and you might not understand all of the details the first time you read through them. However, if you plan to conduct applied empirical work that involves causal inference, you should revisit these again and again as reference.

Optional Books

- *The following books are optional but may prove useful for additional coverage of some of the course topics.*

- *Reference Book for Panel Methods*

- *Causal Inference*

- *Matching*

Preliminary Schedule

The following is a preliminary schedule of course topics. Notice that required readings are marked with a (⋆).

1 Introduction

- Overview, Course Requirements, Course Outline

2 Review of Statistical Concepts Useful for Causal Inference

- Random Variables, Measures of Location and Dispersion
- Inference and Properties of Estimators
- Conditional mean function

3 The Potential Outcome Model

- Counterfactual Responses and the Fundamental Identification Problem
- Estimands and Assignment Mechanisms
- Heterogeneity and Selection

Readings

- Morgan and Winship: Chapter 1-2. (⋆)
- Angrist and Pischke: Chapter 1. (⋆)
4 Randomized Experiments

- Identification of Causal Effects under Randomization
- Implementation, Estimation, Diagnostics, Blocking
- Threats Validity

Readings: Theory of Experiments

- Angrist and Pischke: Chapter 2. (∗)

Readings: Application of Experiments


Readings: Application of Natural Experiments


Readings: Experiments Review Articles

5 Causal Effects under Selection on Observables

5.1 Selection on Observables

- Identification under Selection on Observables
- Subclassification

Readings

- Morgan and Winship: Chapter 3. (⋆)
5.2 Matching Methods

- Covariate Matching, Balance Checks, Properties of Matching Estimators

Readings: Matching Theory

- Morgan and Winship: Chapter 4. (∗)
- Rubin. 2006. Chapters 3 to 5.

Readings: Matching Applications


5.3 Propensity Score Methods

- Identification, Propensity Score Estimation, Matching on the Propensity Score, Weighting on the Propensity Score, Reweighting methods

Readings: Propensity Score Methods Theory

- Morgan and Winship: Chapter 3. (∗)

Readings: Propensity Score Methods Applications


5.4 Regression

• Non-parametric Regression, Identification with Regression

Readings

• Angrist and Pischke: Chapter 3. (⋆)

• Morgan and Winship: Chapter 5. (⋆)

• Chapter in Winship and Morgan on Matching vs Regression.


5.5 Conclusion: Selection on Observables

• Can Non-Experimental Method Recover Causal Effects?

Readings: Comparison of Experimental and Non-experimental Methods


6 Causal Effects under Selection on Time-Invariant Characteristics

6.1 Panel Data Methods

- Fixed Effects and Random Effects Estimation

*Readings: Panel Methods Theory*

- Angrist and Pischke: Chapter 5.1 (*)
- Angrist and Pischke: Chapter 8 (*)

*Readings: Panel Methods Applications*


6.2 Difference-in-Differences Estimators

- Identification, Estimation, Falsification tests

*Readings: DID Theory*

- Angrist and Pischke: Chapter 5.2-5.4 (*)

*Readings: DID Applications*


6.3 Synthetic Control Methods

*Readings*

7 Causal Effects under Selection on Time-variant Characteristics

7.1 Instrumental Variables

- Identification: Using Exogenous Variation in Treatment Intake Given by Instruments
- Imperfect Compliance in Randomized Studies
- Wald Estimator, Local Average Treatment Effects, 2SLS

Readings: Instrumental Variable Theory

- Angrist and Pischke: Chapter 4 (⋆)
- Morgan and Winship: Chapter 7
- Angrist, Joshua D., and Alan B. Krueger. 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments

Readings: Instrumental Variable Critique


Readings: Instrumental Variable Applications

- Angrist and Krueger. 2001 Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments
7.2 The Regression Discontinuity Design

- Sharp and Fuzzy Designs, Identification, Estimation, Falsification Checks

Readings: RDD Theory

- Angrist and Pischke: Chapter 6 (∗)

Readings: RDD Applications


7.3 Sensitivity Analysis

- Nonparametric Bounds
- Formal sensitivity tests

Readings

- Morgan and Winship: Chapter 6 (∗)
8 Distributional Effects

8.1 Quantile Regression

Readings

- Angrist and Pischke: Chapter 7 (∗)

8.2 Distributional Effects in Difference-in-Differences

Readings