Poli 312: Intermediate Quantitative Political Science

Professor Elissa Berwick

Fall 2022

E-mail: elissa.berwick@mcgill.ca
Office Hours: W 1:00-4:00 p.m. (Zoom)
Office: 3610 McTavish, Room 33-3

Web: mycourses2.mcgill.ca
Class Hours: TR 8:35 - 9:55 a.m.
Class Room: ARTS 260

TA: Danielle Bohonos
E-mail: danielle.bohonos2@mail.mcgill.ca

Overview and Goals

This course is designed to continue the introduction to data driven quantitative political analysis begun in Poli 311. The course employs examples from across political science sub-disciplines and is generally relevant to all social science research. Some of the topics covered, particularly in the second half of the course, will be entirely new. However, many other elements of the course will revisit material already introduced in Poli 311, but from a more sophisticated perspective.

The course will meet twice a week for 80 minute sessions. If possible, **please bring laptop computers to class**, for use in classroom activities. I will be available for office hours to answer questions on zoom as indicated on the schedule. Students also are welcome to make appointments for individual zoom meetings to address more specific questions.

Who is this course for?

- Students who took Poli 311 and are eager to enhance their quantitative skills
- Students who may be interested in pursuing a graduate degree in social science and/or a career in data science
- Note that Poli 311 is a prerequisite for this course. If you did NOT take Poli 311, you may only enroll if you provide evidence of equivalent past coursework

Objectives

- Gain a deeper understanding of quantitative empirical research in political science, touching on all three tasks of data analysis: description, prediction, and causation
- Improve skills in `R`

Course Materials

Textbooks

The first required textbook for this course is the same as in Poli 311:

Imai, Kosuke and Lori D. Bougher (2021). *Quantitative Social Science*. Princeton University Press.

We will be focusing on the last three chapters of the Imai book, and taking a much more in-depth look at that material than there was time for in 311.

In the latter half of the course we will also draw on a second textbook, which specifically aims to introduce undergraduates to topics in causal inference:

Angrist, Joshua D. and Jörn-Steffen Pischke (2014). *Mastering 'Metrics*. Princeton University Press.

An additional, optional textbook that is available free online is:

Grolemund, Garrett and Hadley Wickham (2016). *R for Data Science*. http://r4ds.had.co.nz/.

This book is a great resource for coding in `R`. Note that none of the readings in this course are truly "required"; all of the textbooks are there as resources to facilitate your learning, and should be used (or not used) accordingly.

Software

In this course, we will continue using R, a FREE open source language used by data scientists and statisticians across the world. `R` consists of a base environment for data manipulation, calculation and graphical display as well as numerous user-made packages that bundle together more specialized functions.

We will also be using a FREE integrated development environment (IDE) for `R` called RStudio that makes learning and exploring `R` easier. While the learning curve in `R` is steeper than in more expensive programs (such as `Stata` and `SPSS`), there is much more you can do with it!

There are many free online tutorials for downloading and installing `R` and `RStudio`. The `RStudio` team also makes great "cheatsheets" for using their interface (see here) as well as other `R` packages.

Requirements

Problem sets (60%)

There will be 5 problem sets, each worth 12% of the final grade. You will have approximately two weeks to complete each problem set. The first problem set will be due on or around **September 27th**. Each problem set will be due at the start of class time.

Problem set submission

Problem sets must be submitted as PDFs via *MyCourses*. To complete your problem sets, you must use a math-friendly typesetting program such as the `R` variant `rmarkdown` (`.rmd`). Do not submit handwritten assignments or type equations directly into a document editor such as Microsoft Word. You should upload any supporting files (such as raw `R` or `rmd` code) as well.

Collaboration

Collaboration is part of learning how to code. I encourage you to collaborate! But you will not learn how to do statistical programming if you DO NOT write your own code. Please feel free to collaborate with colleagues, but please DO NOT copy each others' code verbatim. You must also write your OWN interpretations of the results.

Final project

Writing a problem set (30%)

The final project for this class will be based on writing your own problem set question(s)! This can be done individually or as part of a **group of up to three**.

Your question(s) should require data analysis based on publicly available statistical data, such as the surveys run by the Canadian Elections Study, World Values Survey, or European Social Survey. A full grading rubric for this assignment will be distributed separately, but as you begin to think about designing your question, you should plan on having it include the following components: (1) a task that requires DESCRIBING the data; (2) a task that requires making PREDICTIONS based on the data; (3) a task that asks for a statistical INFERENCE about the predictions; and (4) a task that entails considering the LIMITATIONS of the data or the analysis.

As part of the assignment, you are also responsible for producing an answer key, containing both code and interpretations of results. You will be assessed on the functionality and accuracy of your code and the precision of your interpretations.

If you are one person, you are responsible for one question; if two people, two questions, etc. If you are part of a group, the questions can all share the same data source, but there must be 2-3 (depending on number of group members) separate description, prediction, inference, and limitation tasks.

Problem set workshops (5%)

In the final week of the course, you will have an opportunity to try out your question(s) on other students during class time. Draft questions and the cleaned data needed to complete them must be posted to `MyCourses` before the start of class on **November 29th**.

Based on discussions with other students, you can then edit question(s) accordingly before the final submission deadline on **December 8th**. Participation in the workshops is **MANDATORY** (worth 5% of final grade).

Attendance (5%)

This course will include time for in-class `R` exercises. By not attending class, students miss out on the opportunity to practice and ask questions about coding and concepts. Attendance will therefore count toward the final grade for the course. Perfect attendance is not required in order to get full credit on this component of the course, but missing 25% or more of the post-add/drop class time without a documented personal or medical reason will be penalized.

Policies

Extraordinary Circumstances Statement

In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change.

Extensions

Each student will start the term with **ten (10) extension days**. You may use an extension day whenever you like and for whatever reason, no explanation required. When you submit an assignment late using your extension days, please indicate at the top of the assignment how many

days you have used on that particular assignment and how many you have left. We will check these at the end of the term. Assignments submitted using days that have already been used up will receive the lowest passing grade. Extensions are good for 24 hours, there are no partial days.

If and when you use all ten days, **no additional extensions will be provided** without a documented *emergency* medical or family reason. If you are unable to complete a homework assignment for documented emergency medical or family reasons, an alternative submission will be arranged. If you cannot provide a valid reason for failing to submit an assignment on time and have used all of your extension days, but still manage to submit the assignment before the **last day of class** you will receive the lowest passing grade.

Re-Grading

Students who wish to contest a grade for an assignment or exam must do so in writing (by email, sent to me) providing the reasoning behind their challenge to the grade received within two weeks of the day on which the assignments are returned. I will re-evaluate the assignment, but also reserve the right to **raise or lower the grade**. Please also see (http://www.mcgill.ca/politicalscience/files/politicalscience/assessment_and_re-read_policy_final.pdf).

Class Discussion Board

I will set up a class discussion board on `MyCourses`. I encourage you to use this to ask questions you may have. *NEVER* post your code or answer to specific homework questions on the discussion board. Please post general questions! If you post homework code on the website, it will be taken down and your grade may be lowered.

Remote Instruction

If you are unable to attend class due to illness, required quarantine, or religious obligation, you must **EMAIL ME** at least **ONE HOUR** before class starts and I will set up Zoom so you can watch the class remotely (preferred) or, if necessary, as a recording within 48 hours.

The purpose of the Zoom option is to make up for unavoidable absences, not to allow for fully remote instruction.

Copyright of Lectures

All slides, video recordings, lecture notes, etc. remain the instructor's intellectual property. As such, you may use these only for your own learning (and research, with proper referencing/citation) ends. You are not permitted to disseminate or share these materials; doing so may violate the instructor's intellectual property rights and could be cause for disciplinary action.

I remind everyone of their responsibility in ensuring that this video and associated material are not reproduced or placed in the public domain. This means that each of you can use it for your educational (and research) purposes, but you cannot allow others to use it, by putting it up on the Internet or by giving it or selling it to others who may also copy it and make it available. Please refer to McGill's Guidelines for Instructors and Students on Remote Teaching and Learning for further information.

Academic Integrity

Course Policy on Computer Code

As discussed in the problems set section, verbatim copying other people's computer code constitutes plagiarism. Moreover, data programming is learned through trial and error. *Please do not under any circumstances copy another students code.* You may of course collaborate with colleagues, but please write your own code! If you are found to have plagiarized, you may be referred to the appropriate Dean. The instructors reserve the right to use software to compare the code that has been written by different students.

McGill Policy

"McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures" (see www.mcgill.ca/students/srr/honest/ for more information).

Language of Submission

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Conformément à la Charte des droits de l'étudiant de l'Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté (sauf dans le cas des cours dont l'un des objets est la maîtrise d'une langue).

Disabilities Policy

As the instructor of this course I endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with me and the Office for Students with Disabilities, 514-398-6009.

End of Course Evaluations

End-of-course evaluations are one of the ways that McGill works towards maintaining and improving the quality of courses and the student's learning experience. You will be notified by e-mail when the evaluations are available. Please note that a minimum number of responses must be received for results to be available to students.

Class Schedule

Week 01, 08/29 - 09/02: Introduction

Week 02, 09/05 - 09/09: (Reviewing R; Probability) Dealing with data in R; Probability

Topics: the `tidyverse`, `ggplot2`, reshaping data in R, probability and simulation, functions and iterations in R

Resources

- Grolemund & Wickham Chapters 3, 5, 19, 21
- Data manipulation lectures
- ggplot lectures
- Imai 6.1

Week 03, 09/12 - 09/16: (Probability) Random variables and their distributions

No class Thursday, September 15th

Topics: random variables, measures of location and dispersion, probability distributions.

Resources

- Imai, Chp 2.6, 6.2 - 6.4

Week 04, 09/19 - 09/23: (Probability/Inference) Multiple random variables; Point estimation

Topics: joint, conditional and marginal distributions, measures of association, properties of estimators, point estimation, sampling distribution, standard error, large sample theorems.

Resources

- Imai, Chp 3.6, 7.1

Week 05, 09/26 - 09/30: (Inference) Interval estimation and hypothesis testing

Topics: confidence intervals, hypothesis testing, t-distribution, p-values.

Resources

- Imai Chp 4.1 - 4.2, 7.1 - 7.2

Week 06, 10/03 - 10/07: (Reviewing regression) Regression and its properties

No office hours Wednesday, October 5th.

Topics: non-parametric regression, linear regression model, properties of OLS, homoskedasticity

Resources

- Imai, Chp 4.1 - 4.2, 7.3

Week 07, 10/10 - 10/14: (Reviewing regression) Inference for regression

No class Tuesday, October 11th. Friday, October 14th follows a Tuesday schedule.

Topics: confidence intervals and hypothesis tests for coefficients, testing workflow.

Resources

- Imai, Chp 4.1 - 4.2, 7.3

Week 08, 10/17 - 10/21: (Refining regression) Multiple regression and its pitfalls

Topics: multiple regression, mechanics and inference with multiple regressors, marginal effects, omitted variables bias, multicollinearity.

Resouces

- Imai, Chp 4.1 - 4.2, 7.3
- Bailey, Chapter 5 (scan)

Week 09, 10/24 - 10/28: (Refining regression) Dummy variables and interaction

Topics: dummy variables, interaction terms, polynomials, marginal effects

Resources

- Bailey Chp 6 - 7 (scan)

Week 10, 10/31 - 11/04: (Refining regression; Causation with regression) Prediction and uncertainty; Potential outcomes

Topics: making predictions and reporting their results, confidence intervals for predicted values and marginal effects, potential outcomes model,

Resources

- Angrist & Pischke Chp 1
- Imai 4.1 - 4.2, 2.3

Week 11, 11/07 - 11/11: (Causation with regression) Randomized experiments

Topics: randomized experiments, heterogeneous treatment effects, inference for experiments, threats to validity, covariate adjustment, non-standard standard errors

Resources

- Angrist & Pischke Chp 2
- Imai Chp 2.4, 4.3 (review)

Week 12, 11/14 - 11/18: (Natural experiments) Regression discontinuity designs

Resources

- Angrist & Pischke Chp 4
- Imai Chp 4.3.4

Week 13, 11/21 - 11/25: (Natural experiments) Differences-in-differences

Resources

- Angrist & Pischke Chp 5
- Imai Chp 2.5

Week 14, 11/28 - 12/02: Problem set workshops

Final project due December 8th