Poli 618: Quantitative Analysis

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: MW 10:05-11:25
Class Room: ARTS 230

TA: Maxime Blanchard
E-mail: maxime.blanchard@mail.mcgill.ca

Overview and Goals

This course is designed to introduce graduate students to the exciting world of data driven quantitative political analysis. The course employs examples from across political science sub-disciplines and is generally relevant to all social science research.

One hour lab sections will meet weekly with the TA at a time to be arranged. Labs will focus on practical implementations of course material and occasionally reviewing mathematical background concepts.

I will be available for open 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?

- This is your first semester-long graduate quantitative methods coursewith a focus on data analysis
- You are willing to spend time considerable outside of the classroom tolearn the course materials, as data analysis is a skill learned by doing
- You may be interested in taking higher level statistical classes focused on causal inference (POLI 666) and prediction (POLI 667) in the future

Objectives

- Learn the basic tools of quantitative empirical research in political science
- Obtain skills in R, a highly powerful and FREE programming language used extensively by academics in political science across the world, as well as the open source and data science community
- Enhance quantitative literacy and understand scientific replicability

Course Materials

Textbooks

The textbooks used in this course should be viewed as resources that will reinforce learning from the lectures and provide additional information on certain topics. The lectures will not directly mirror the textbooks, and the readings are not a replacement for the lectures.

Given people's various backgrounds, we will have one required and several optional textbooks in addition to the assigned articles. The Bailey book is a great book and very applied. Everyone should read it. For those of you seriously interested in pursuing quantitative analysis, you should then read the Fox book on top of the Bailey book. The Moore book should be used for additional math review as needed. The Grolemund and Wickham book is a great tool (that is online and free) and can often be used in lieu of videos or to help with coding

Bailey, Michael A. (2020). *Real Stats: Using Econometrics for Political Science and Public Policy*. Oxford University Press, USA.

Fox, John (2015). *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, Incorporated.

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

Moore, Will H. and David A. Siegel (2013). *A Mathematics Course for Political and Social Research*. Princeton University Press.

Software

In this course, we will be using the statistical computing environment 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 (50%)

There will be 5 problem sets, each worth 10% of the final grade. These assignments will be due approximately every two weeks, at the **start of class time**, with extra time over the fall break. The first problem set will be due on or around **September 28th**. Note that the final problem set will be due during the presentation weeks, there is no gap time between when problem sets end and when presentations begin. Plan accordingly!

Problem set submission

Problem sets must be submitted via `GitHub`. To complete your homework, you must use an R variant, the GitHub compatible `.Rmd` (otherwise known as `rmarkdown`) files. You should submit

both the `.Rmd` file and the PDF it produces, organized within a folder in your `GitHub` repository. Do not submit your problem set using Microsoft Word or any other document editor. It will not be graded.

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

The final project will be based on quantitative data analysis, and can be *either* a replication paper with an **extension** of the original paper *or* an **original** paper of your own.

Replication papers should not be based on simply duplicating the original results, but should instead either (1) ask a new question of the same data; (2) highlight where the replicated paper's analysis could have been improved; or (3) test alternative explanations of the replicated paper's argument.

Meanwhile, an original paper must contain (1) a clear theory proposing a relationship between explanatory variable(s) on an outcome variable; (2) use of linear regression (or some other model cleared with professor); and (3) a clear discussion of both findings and limitations of the paper.

The grade for the final project is based on an initial proposal, a presentation of results, and a paper writing up those results. You MAY work with a partner on this project, but expectations will increase accordingly.

If you are working with a partner, you must inform the instructor by **September 19th**.

Project proposal (10%)

Due **October 21st**. This should be a one page write up in `rmarkdown` explaining the data set/s which you are going to use and the question you will ask. You should also highlight your outcome variable. Submission will be via `Github`.

Presentation (10%)

Each student (or pair of students) will present their idea for the final project and any initial results to the class during the **final 3 weeks** of the term. Presentations should last approximately ten minutes, to be followed by five minutes of questions.

The presentation grade will be based both on the presentation AND on engagement with the work of classmates. (Tip: you can also use `rmarkdown` or `quarto` via `RStudio` to generate presentations, but you are not required to do so!)

Paper (30%)

Due **December 9th**. Students will submit a final paper of article length via `GitHub`, alongside all code used in the paper. Students may use a chapter of their master or Ph.D. thesis as a research paper.

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. Each extension day is 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 date 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.

Note that the extension days CANNOT be combined with the end of semester TA extension

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

We will use `GitHub` for class discussions. Once you join the team `fall2022_students` you will have access to the discussions page. 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 to quantitative data and statistical computing

Lecture topics: Introduction and course outline; causation and prediction; observational and experimental data; R, RStudio and GitHub

Resources

- Andy Matushak, Why Books Don't Work.
- Grolemund & Wickham, Chapter 27
- Intro R lectures Chapters 1 (Intro to Basics), 2 (Vectors), 4 (Factors), and 5 (Data Frames)
- R-markdown lectures
- Moore, Chapter 1
- Bailey, Chapter 1 - 2

Tasks

- Schedule lab sections

Week 02, 09/05 - 09/09: Probability and simulation

No class Monday, September 5th

Lecture topics: Probability; simulated solutions to probability questions. Lab: data manipulation via dplyr.

Resources

- Data manipulation lectures

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

Lecture topics: Discrete and continuous random variables; probability distributions; measures of location and dispersion; joint and conditional distributions; conditional expectation. Lab: distributions.

Resources

- Moore, Chapter 9
- Wooldridge, Appendix B.1, B.3, B.5, C.5 (scan)
- Grolemund & Wickham Chapters 3, 7
- Bailey, Appendix B - C, F - I
- Imai 6.3 - 6.4 (scan)

Week 04, 09/19 - 09/23: Estimation and inference

Lecture topics: Properties of estimators; sampling distributions; large sample theorems; point estimation; interval estimation; confidence intervals; t-distribution. Lab: data visualization via ggplot.

Resources

- Imai 7.1 (scan)
- Wooldridge, Appendix C.1 - C.3 (scan)
- ggplot lectures

Tasks

- Decide on group for final project by **September 19th**

Week 05, 09/26 - 09/30: Hypothesis testing

No class Monday, September 26th.

Lecture topics: Statistical testing; p-values. No lab.

Resources

- Bailey, Appendix D - E, Chapter 4
- Imai 7.2 (scan)
- Wooldridge, Appendix B.2 & B.4 & C.6 (scan)

Week 06, 10/03 - 10/07: Introduction to regression (recorded)

No in-person class Monday, October 3rd.

No class or office hours Wednesday, October 5th.

Lecture topics: Non-parametric regression, bias-variance trade-off, local linear regression. Lab: calculus review

Week 07, 10/10 - 10/14: Simple linear regression

No class October 10th or 12th; Thursday, October 13th is a Monday make-up class

Lecture topics: Principle of least-squares; analysis of variance for regression; properties of OLS. No lab.

Week 08, 10/17 - 10/21: Inference for regression and regression with two regressors

Lecture topics: Homoskedasticity; Gauss-Markov assumptions; confidence intervals and hypothesis tests for coefficients; testing workflow; large sample inference; mechanics and inference with two regressors; comparing models. Lab: regression.

Resources

- Moore, Chapter 2 (if you need algebra review)
- Moore, Chapters 5-6 (if you need calculus review)
- Bailey, Chapter 3 - 5
- Fox, Chapter 5.1 - 5.2, 6.1 - 6.2

Tasks

- Submit project proposal by **October 21st**

Week 09, 10/24 - 10/28: Bias, dummy variables and interactions

Lecture topics: Omitted variables bias; multicollinearity; dummy variables; interaction terms; polynomials, prediction. No lab.

Resources

- Bailey, Chapter 5 - 7
- Fox, Chapter 7

Berry, William D., Matt Golder, and Daniel Milton (2012). "Improving Tests of Theories Positing Interaction". In: *The Journal of Politics* 74.3, pp. 653-671.

Brambor, Thomas, William Roberts Clark, and Matt Golder (2006). "Understanding Interaction Models: Improving Empirical Analyses". In: *Political Analysis* 14.1, pp. 63-82.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu (2019). "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice". In: *Political Analysis* 27.2, pp. 163-192.

Week 10, 10/31 - 11/04: Matrix regression and its properties

Lecture topics: Regression and prediction in matrix form; assumptions in matrix form; testing multiple hypotheses. Lab: simulation

Resources

- Moore, Chapter 12 (if you need matrix review)
- Fox, Chapter 9.1 - 9.5

Week 11, 11/07 - 11/11: Linear regression diagnostics and fixes

Lecture topics: Residuals; outliers, leverage and influence; heteroskedasticity; robust standard errors; autocorrelation. Lab: biases

Resources

- Fox, Chapters 11 - 13

Tasks

- Sign up for a presentation date

Week 12, 11/14 - 11/18: Regression extensions

Lecture topics: Non-linearity; linear probability model; logit and probit functions; GLMs; interpretion and uncertainty in logit models. Lab: logistic

Resources

- Bailey Chapter 12
- Fox Chapter 14

Week 13, 11/21 - 11/25: Regression extensions continued; Presentations

Lecture topics: Predicting from logit models.

Resources

- Bailey Chapter 12
- Fox Chapter 14

Week 14, 11/28 - 12/02: Presentations

Week 15, 12/05 - 12/09: Presentations

Last class Monday, December 5th

Tasks

- Submit final paper by **December 9th**

Other recommended textbooks

There are many other important textbooks and at some point you may find yourself looking for a different explanation of something you didn't understand – or looking to go deeper. Here are some places to start.

Angrist, Joshua D. and Jorn-Steffen Pischke (2008). *Mostly Harmless Econometrics*. Princeton University Press.

Gailmard, Sean (2014). *Statistical Modeling and Inference for Social Science*. Cambridge University Press.

Gelman, Andrew, Jennifer Hill, and Aki Vehtari (2021). *Regression and other stories*.

Greene, William H. (2012). *Econometric Analysis*. Pearson College Division.

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

Wooldridge, Jeffrey M. (2019). *Introductory Econometrics: A Modern Approach*. Cengage Learning.