Poli 618: Quantitative Analysis

Fall 2024

Instructor

Professor Elissa Berwick
elissa.berwick@mcgill.ca
Wednesday 10:00-12:00
3610 McTavish, Room 33-3

Course

Monday, Wednesday
1:05-2:25 PM
LEA 424
mycourses2.mcgill.ca

Overview

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.

The course will combine traditional lectures with in-class coding exercises. Whenever possible, **please bring laptop computers to class** (fully charged!), for use in classroom activities.

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

I will be available for office hours to answer questions as indicated on the schedule. Students also are welcome to make appointments for individual meetings to address more specific questions.

Who is this course for?

- This is your first semester-long graduate quantitative methods course with a focus on data analysis
- You are willing to spend time considerable outside of the classroom to learn the course materials, as data analysis is a skill learned by doing
- You may be interested in taking higher level statistical classes 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

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. 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.

Additional scanned readings will be posted to `MyCourses`.

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.

Software

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 ma-

nipulation, 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.

Evaluation

Lab reports (50%)

Over the course of the semester, you will be responsible for submitting five lab reports, each worth 10% of the final grade. These reports will be due approximately every two weeks, according to the following schedule:

- Report 1: weeks 2-4, due end of week 5
- Report 2: weeks 5-6, due end of week 7
- Report 3: weeks 7-8, due end of week 9
- Report 4: weeks 9-10, due end of week 11
- Report 5: weeks 11-12, due end of week 13

Each lab report will consist of a set of data analysis tasks based on methods introduced in class during the previous 2-3 weeks. You will also practice these methods during your weekly lab sections. For the first two lab reports, you will be given a dataset by the instructor and asked to write up a report that applies what you've learned. For the final three lab reports, you will instead apply the data analysis tasks to the dataset that you have selected for your final project (see below).

Grades for lab reports will be based on correctly applying and interpreting results from methods learned in class as well as the thoroughness of responses. Reports must be submitted in PDF format and should look as professional as possible, with all figures and tables properly labelled. To make your reports, you may either insert output from `R` output into a typesetting program such as LaTeX, or `Typst`, or you may use a notebook program such as `rmarkdown` or `quarto` that directly incorporates `R` code. Instructors will provide support for using `rmarkdown`.

You must submit both your code files (`.R`, `.Rmd`, or `.qmd`) and the final PDF. Do not submit your reports using Microsoft Word or any other document editor. They will not be graded.

You can consult any online resources as you see fit, but if you use AI assistance you **must** submit chat logs along with your work. The instructor or TA may ask you to walk through submitted lab reports in person if there is reason to believe this policy is being violated.

Final project (50%)

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

Replication papers should not be based on simply duplicating the source paper's 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.

Project proposal (10%)

Due **October 7th**. This should be a one page write up explaining the data set/s which you are going to use and the question you will ask. You should also highlight your outcome variable. Grades will be based on clearly establishing a link between the data and the question.

Presentation (10%)

Each student 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 11th**. Students will submit a final paper of at least ten pages, alongside all code used in the paper. Students may use a chapter of their MA or Ph.D. thesis as a research paper. Grades will be based on the correct application of methods covered in the course, as well as how well the paper uses quantitative methods to test or advance a theory.

Schedule

Part 1: Fundamentals of statistics

Week 1: Introduction; data and computing, 08/28

- Introduction and course outline; causation and prediction; observational and experimental data; `R`, `RStudio`
- Grolemund & Wickham, Chapter 27
- Intro R lectures Chapters 1 (Intro to Basics), 2 (Vectors), 4 (Factors), and 5 (Data Frames)
- R-markdown lectures
- ggplot lectures
- Moore, Chapter 1

- Bailey, Chapter 1 - 2

Week 2: Random variables, 09/02-09/04

- Discrete and continuous random variables; probability distributions; measures of location and dispersion; joint and conditional distributions; conditional expectation
- 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 3: Point and interval estimation, 09/09-09/11

- Properties of estimators; sampling distributions; large sample theorems; point estimation; interval estimation; confidence intervals; t-distribution
- Imai 7.1 (scan)
- Wooldridge, Appendix C.1 - C.3 (scan)

Week 4: Hypothesis testing and simulation 09/16-09/18

- Statistical testing; p-values; simulation from distributions
- Bailey, Appendix D - E, Chapter 4
- Imai 7.2 (scan)
- Wooldridge, Appendix B.2 & B.4 & C.6 (scan)

Part 2: Regression

Week 5: Simple linear regression, 09/23-09/25

- Principle of least-squares; analysis of variance for regression; properties of OLS; homoskedasticity; Gauss-Markov assumptions; confidence intervals and hypothesis tests for coefficients
- Moore, Chapter 2 (if you need algebra review)
- Moore, Chapters 5-6 (if you need calculus review)
- Bailey, Chapter 3-4
- Fox, Chapter 5.1

Week 6: Multiple regression, 09/30-10/02

- Mechanics and inference with two regressors; comparing models; omitted variables bias; multicollinearity
- Bailey, Chapter 4-5
- Fox, Chapter 5.2

Week 7: Indicators and interactions, 10/07-10/09

- Dummy variables; interaction terms
- Bailey, Chapter 5-6

- Fox, Chapter 7

Reading break, 10/14-10/18

Week 8: Transformations and predictions 10/21-10/23

- Polynomials, logs, marginal effects, average effects, prediction and uncertainty
- Fox, Chapter 4
- Gelman, Andrew, Jennifer Hill, and Aki Vehtari (2020). *Regression and Other Stories*. Cambridge University Press. Chapter 12 (scan)
- Arel-Bundock et al, "How to Interpret Statistical Models Using `marginaleffects` for R and Python", https://marginaleffects.com/files/marginaleffects_arel-bundock_greifer_heiss_jss5115.pdf

Part 3: Advanced regression

Week 9: Matrix regression 10/28-10/30

- Regression and prediction in matrix form; assumptions in matrix form; testing multiple hypotheses
- Moore, Chapter 12 (if you need matrix review)
- Fox, Chapter 9.1 - 9.5

Week 10: Diagnostics and fixes, 11/04-11/06

- Residuals; outliers, leverage and influence; heteroskedasticity; robust standard errors; autocorrelation
- Fox, Chapters 11 - 13

Week 11: Logistic regression, 11/11-11/13

- Non-linearity; linear probability model; logit and probit functions
- Bailey Chapter 12
- Fox Chapter 14

Week 12: Using non-linear models; Presentations, 11/18-11/20

- Interpretion and uncertainty in logit models; generalized linear models
- Bailey Chapter 12
- Fox Chapter 14

Week 13: Presentations, 11/25-11/27

Week 14: Presentations, 12/02-12/04

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 `MyCourses` for class discussions. I encourage you to use the discussion board 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

Verbatim copying of 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.