

POLI 618: Quantitative Analysis

Professor Elissa Berwick

Autumn 2020

Instructor: **Elissa Berwick**
elissa.berwick@mcgill.ca
3610 McTavish, Room 26-4

Fixed lectures: T 11:35-12:35
Flexible lectures: posted on Thursdays
Office hours: T, Th 12:35 - 1:15

TA: **Olivier Bergeron-Boutin**
olivier.bergeron-boutin@mail.mcgill.ca

Lab Hours:
TBD

1 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.

The course is divided into fixed lectures, flexible lectures, and labs. The first lecture of term, on Thursday, September 3rd, will be held live for as many students as possible, to facilitate introductions. In subsequent weeks, a fixed zoom lecture will be held on Tuesdays, during the first hour of the scheduled class time. Each fixed lecture will introduce the concepts covered for the week, and students will have the chance to stop the lecture and ask questions in real time. All students who can attend are encouraged to do so, but these sessions will also be recorded for students who are unable to attend.

A second flexible lecture, of approximately forty minutes, will be posted online on Thursdays for students to watch before the next fixed class session. These flexible lectures will wrap up the week's material. The Thursday lectures will be accompanied by interactive R notebooks or other comprehension exercises that students will complete after watching the lecture. Following the flexible lectures, students are particularly encouraged to submit questions to the online discussion board to be discussed by their peers and/or in the next session.

I will be available for optional, public office hours to answer questions on zoom immediately following the fixed lecture on Tuesdays and at the same time on most (but not all)

Thursdays, as indicated on the schedule. Students are also welcome to make appointments for individual zoom meetings to address more specific questions.

Lab sections of approximately forty minutes will meet weekly with the TA at a time to be arranged, with the possibility of multiple time slots to accommodate all students. Labs will focus on practical implementations of course material and occasionally reviewing mathematical background concepts.

Who this course is 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 focused on causal inference (POLI 666) and prediction (POLI 667) in the future.

Objectives

1. Learn the basic tools of quantitative empirical research in political science.
2. 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.
3. Enhance quantitative literacy and understand scientific replicability.

2 Course Materials

Required Textbooks

Given people's various backgrounds, we will have three 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. Unless otherwise cleared with me, everyone is required to do the math review problems assigned from the Moore book. The Golemund 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. (2015). *Real Stats: Using Econometrics for Political Science and Public Policy*. 1st. New York: Oxford University Press.

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

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

Recommended Textbooks

Fox, John (2015). *Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Los Angeles: SAGE Publications, Inc. ISBN: 978-1-4522-0566-3.

Required 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.

3 Course Requirements

There are 1000 points available in the class. Therefore, for each 10% of the grade, 100 points are allotted.

- **40%** *Problem sets*. A quarter of this grade will be from peer-reviewing each others' assignments on *GitHub*.
- **40%** *Final paper*. A paper based on quantitative data analysis — likely a replication exercise or part of your dissertation or MA Thesis.
- **20%** *Poster session*. Based on the presentation of a conference-style poster and participation in a virtual poster session.

Problem sets

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. Do not submit your homework using Microsoft Word or any other document editor. It will not be graded. Part of your homework grade will come from peer-reviewing your colleagues' homework assignments via the course *GitHub*. There will be approximately 5 problem sets. I reserve the right to lower the number of problem sets if needed. In principle, these assignments will be due every 2 weeks starting week 2 for the subsequent 12 weeks of the term. Collaboration is part of learning how to code. I encourage you to collaborate! But you do 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.

Interim Data Set and Check-in

All students must submit a one page write up of the data set they are going to use and the research question they are going to ask by **October 22nd**. 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*.

Final paper

All students will submit a final paper that is of article length. This will be done via *Github*. This can be *either* a replication paper with an extension of the original paper *or* a new paper. It is highly recommend you do a replication. Please come talk me as early in the term as possible if you want to write an original paper. 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); 3) a clear discussion of both findings and limitations of the paper. Students may use a chapter of their master or Ph.D. thesis as a research paper. The paper is due by **December 18th**.

Poster sessions

Each student will also make a conference-style poster summarizing their research and answer questions about it from their classmates during a virtual poster session. Posters should be circulated among all students in PDF format by **November 22nd**. The poster sessions will take place during the last two weeks of classes (weeks of November 23rd and 30th), and if necessary, students will be assigned to different sessions based on their time zones. The grade for the poster session will be based both on the student's own poster and on engagement with the work of classmates.

A poster should follow the structure of your paper, and thus it is a helpful way to think about the organization of your paper before writing it. Some suggestions:

1. *Use keywords and bullet points*: Don't use full sentences—your audience will never read them. Try to use keywords (or half sentences when needed), and make sure that you use only one line to deliver each point.
2. *Choose key figures*: Select just a few key figures that best convey the argument you are making, and center the poster around them.
3. *Use a template*: There are many online templates to help you make posters easily, either in L^AT_EX (see <https://github.com/deselaers/latex-beamerposter>) or in RMarkdown (see <https://github.com/brentthorne/posterdown>).

4 Class 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.

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 paper, 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).

Make-Up Work

If you are unable to complete a homework assignment for documented emergency medical or family reasons, an alternative submission date will be arranged. The alternative arrangement is only open to those who can provide a valid medical/family reason. If you cannot provide a valid reason for failing to submit an assignment on time, you will receive zero points for the submission.

Class Discussion List and E-mail

I will set up a class discussion list on *MyCourses*. I encourage you to use this mailing list to ask questions you may have, especially if you are taking lectures asynchronously. *NEVER* post your code or specific homework questions on the course list server. Please post general questions! If you post homework code on the website, it will be taken down and your grade may be lowered.

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.

Recording

By enrolling in a remote course, you accept that fixed sessions will be recorded. You must consent to being recorded if you are attending a lecture or participating in a component of a course that is being recorded. You will be notified through a “pop-up” box in Zoom if a lecture or portion of a class is being recorded. If you are not comfortable being in a class that is recorded, you may decide to not take part by logging off Zoom. Students who log off will be able to later watch the video recording in MyCourses.

For pedagogical reasons and for the enrichment of the experience of all students, attendance may be monitored and/or active participation may be expected or required during fixed (synchronous) class time. As such, you may be asked to turn on your camera and audio. If you do not have the necessary resources (e.g., adequate Internet bandwidth or equipment) to do so, inform your instructor at the beginning of term so that appropriate accommodations can be made.

In addition to the recording of your image and voice, your name (or preferred name) may be displayed on screen, and your instructor may call your name during the lecture. As such, this personal information will be disclosed to classmates, whether during the lecture or in viewing the recording. By remaining in classes that are being recorded, you accept that personal information of this kind may be disclosed to others, whether during the lecture or in viewing the recording.

Recordings will be deleted at the end of the semester to protect students’ privacy.

Netiquette

The University recognizes the importance of maintaining teaching spaces that are respectful and inclusive for all involved. To this end, offensive, violent, or harmful language arising in contexts such as the following may be cause for disciplinary action:

1. Username (use only your legal or preferred name)
2. Visual backgrounds
3. Chat boxes

To maintain a clear and uninterrupted learning space for all, you should keep your microphone muted throughout your class, unless invited by the instructor to speak. You should follow instructors’ directions about the use of the chat function on remote learning platforms, and should NOT use the chat function for private conversations during class time.

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.

5 Class Schedule

** indicates no Thursday office hours*

Week 01, 09/03: Introduction to quantitative inference*

Lecture topics: Introduction and course outline; causation and prediction; internal and external validity. No lab.

- AUDIO: Andy Matushak, [Why Books Don't Work](#).
- READING:
 - Moore, Chapter 1
 - Bailey, Chapter 1
- TASKS
 - Fill out Pre-course Survey
 - Schedule lab sections
 - Complete [tryGit](#)
 - Install git on your computer and push a test file to GitHub

Week 02, 09/08 - 09/10: Quantitative data and statistical computing*

Lecture topics: Observational and experimental data; introduction to R and RMarkdown; discrete and continuous data; missing data; replication and GitHub. Lab: data manipulation via dplyr.

- VIDEOS (for those who want to refresh skills from R-camp):
 - [Intro R lectures](#) Chapters 1 (Intro to Basics), 2 (Vectors), 4 (Factors), and 5 (Data Frames)
 - [R-markdown lectures](#)
 - [Data manipulation lectures](#)
- READING:
 - Golemund & Wickham, Chapter 27
 - Bailey, Chapter 2
 - Dafoe, Allan (2014). “Science Deserves Better: The Imperative to Share Complete Replication Files”. In: *PS: Political Science & Politics* 47.1, pp. 60–66.
 - Eubank, Nicholas (2016). “Embrace your Fallibility: Thoughts on Code Integrity”. In: *The Political Methodologist* 23.2, pp. 10–15. (Visited on Nov. 21, 2016).
 - Moravcsik, Andrew (2010). “Active Citation: A Precondition for Replicable Qualitative Research”. In: *PS: Political Science & Politics* 43.01, pp. 29–35. (Visited on Nov. 30, 2016).

Week 03, 09/15 - 09/17: Describing and visualizing data

Lecture topics: Introduction to graphics; visualizing univariate and bivariate distributions; measures of location, dispersion, and association; clustering. Lab: data visualization via ggplot.

- VIDEOS
 - Data camp: [ggplot lectures](#)
- READING:
 - Imai, Sections 2.6, 3.3, 3.6 - 3.7 (scan)
 - Grolemund & Wickham Chapters 3 and 7
 - Kastellec, Jonathan P. and Eduardo L. Leoni (2007). “Using Graphs Instead of Tables in Political Science”. In: *Perspectives on Politics* 5.4, pp. 755–771. ISSN: 1541-0986, 1537-5927. (Visited on Oct. 25, 2016).
 - King, Gary, Michael Tomz and Jason Wittenberg (2000). “Making the Most of Statistical Analyses: Improving Interpretation and Presentation”. In: *American Journal of Political Science* 44.2, pp. 347–361. ISSN: 0092-5853. (Visited on Apr. 28, 2011).

Week 04, 09/22 - 09/24: Probability and random variables*

Lecture topics: Probability; types of uncertainty; random variables; probability distributions; expectation and variance; large sample theorems. Lab: iteration and distributions.

- READING:
 - Moore, Chapter 9
 - Wooldridge, Appendix B (scan)
 - Imai, Chapter 6 (scan)

Week 05, 09/29 - 10/01: Confidence and significance

Lecture topics: Estimators; point estimation; interval estimation; hypothesis testing for means. Lab: calculus review.

- READING:
 - Imai, sections 7.1-7.2.5 (scan)
 - Wooldridge, Appendix C (scan)

Week 06, 10/06 - 10/08: Linear regression

Lecture topics: Systematic and stochastic variation; sum of squared errors; OLS derivation; uncertainty of coefficients; hypothesis testing for coefficients. Lab: linear regression.

- **READING:**
 - Moore, Chapter 2 (if you need algebra review)
 - Moore, Chapters 5-6 (if you need calculus review)
 - Bailey, Chapters 3-5
 - Fox, Chapter 2

Week 07, 10/13 - 10/15: Interaction and simulation

Lecture topics: interaction terms; marginal effects and predicted values; uncertainty of predicted values; simulation. Lab: simulation.

- **READING:**
 - Moore, Chapter 7
 - Bailey, section 6.4
 - Fox, section 7.3
 - 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. ISSN: 0022-3816, 1468-2508.
 - Brambor, Thomas, William Roberts Clark and Matt Golder (2006). “Understanding Interaction Models: Improving Empirical Analyses”. In: *Political Analysis*, pp. 63–82.

Week 08, 10/20 - 10/22: Transformation and comparison

Lecture topics: Polynomial and logistic transformation; dummy variables; curve fitting; goodness of fit; omitted variable bias; bias-variance tradeoff. Lab: matrix math review.

- **READING:**
 - Moore, Chapter 3
 - Bailey, Chapters 6.1-6.3, 7
 - Fox, Chapters 4, 7.1-7.2
- **PAPER CHECK IN : Thursday, Oct 22**

Week 09, 10/27 - 10/29: Matrix presentation of LS

Lecture topics: Matrix and vector; identity, transpose, and inverse matrices; LS regression in matrix form; inference for multiple linear regression; testing multiple hypotheses. Lab: matrix math in R.

- **READING:**
 - Moore, Chapter 12
 - Fox, Chapter 9.1-9.2

Week 10, 11/03 - 11/05: Assumptions & properties of linear regression

Lecture topics: Variance-covariance matrix; heteroskedasticity; independence and multicollinearity; exogeneity and endogeneity; Gauss-Markov conditions. Lab: biases.

- READING:
 - Bailey, Chapter 14
 - Fox, Chapter 6

Week 11, 11/10 - 11/12: Linear regression diagnostics and fixes

Lecture topics: Plotting residuals; standardized and studentized residuals; outliers, leverage and influence; weighted least-squares; heteroskedasticity robust and cluster robust standard errors; autocorrelation. Lab: biases

- READING:
 - Moore Chapter 12
 - Fox Chapter 9.1-9.2

Week 12, 11/17 - 11/19: Extensions and conclusion*

Lecture topics: Data generating process; non-linear models; binary and count data; logit and probit functions; GLMs in R; linear probability model. Lab: logistic

- READING:
 - Bailey Ch. 12
 - Fox Ch. 14
- POSTER DUE : Sunday, November 22nd

Week 13, 11/24 - 11/26: Virtual poster sessions

Week 14, 12/01: Virtual poster sessions cont.

- PAPER DUE : Friday, December 18th

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 Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

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

Gelman, Andrew and Jennifer Hill (2007). *Data Analysis Using Regression and Multi-level/Hierarchical Models*. New York: Cambridge University Press.

Greene, William H. (2003). *Econometric analysis*. Pearson Education.

Imai, Kosuke. (2017). *Quantitative Social Science: an Introduction*. Princeton, NJ: Princeton University Press.

Wooldridge, Jeffrey. (2009) *Introductory Econometrics*. New York: South-Western. 4th edition.