Quantitative Methods - SOCI504

McGill University, Winter Term 2016

Thursdays 2:30 to 5:30 in Room 1285, 688 Sherbrooke

1 Overview

This course is designed to introduce you to quantitative social science methodology. The primary focus of the course will be on regression models with special attention paid to least squares linear regression and logistic regression models. These models are a good starting point as more complicated models can be thought of as extensions of these models.

Understanding basic regression models is important even if you don't plan to become a quantitative social science researcher. A large share of published work sociology relies on quantitative analysis. Probably two-thirds of articles in a typical ASR issue use some form of regression analysis or related technique. Understanding this work and being able to critique it will be a critical part of being a successful researcher.

Quantitative modeling can be best understood as an endeavor to approximate some social process in a stylized fashion. Our models will never be accurate - the real question is whether a particular model is close enough to reality to allow us to answer the questions we are interested in. Determining whether a particular model is adequate is an art. It takes a lot of practice and experience. For this reason we will be getting our hands dirty with real data as much as possible in this course. But this class can only provide the starting point. To truly master the material you need to gain experience analyzing a lot of data. I can't emphasize this enough. It would be a very good idea for you to spend a large portion of your free time analyzing both simulated and real data so that you gain more practical experience and intuition about theses approaches.

Software:

In this course we will be using the open-source statistical programming language $\mathbb{R}$ which is probably the most widely used statistical software. $\mathbb{R}$ also allows for carrying out state-of-the-art computer-based simulations and generates really nice publication-quality graphics. The software runs under a wide array of operating systems. $\mathbb{R}$ can be downloaded for free at http://www.r-project.org/. Learning $\mathbb{R}$ might seem a bit challenging at first, but you will realize that it is incredibly powerful. The lab sessions will be devoted to learning data analysis techniques in $\mathbb{R}$.

Readings

We will be using the following textbook which should be available at the McGill bookstore. Note: this edition just came out literally over the holiday break

> John Fox. 2016. *Applied Regression and Generalized Linear Models.* Sage.

Though not required I can strongly recommend:

> Fox, John. 2002. *And R and S-Plus Companion to Applied Regression.* Sage.
>
> Achen, Christopher. 1982. *Interpreting and using regression.* Sage

There will also be some emphasis on data visualization and graphical methods for data analysis. Some classic works in that field that I cannot recommend highly enough are:

> Cleveland, William S. 1993. *Visualizing Data.* Summit, NJ: Hobart

In addition I will assign articles and chapters from books.

2 Components of the course and evaluation

This course is scheduled once a week for three hours. We will likely not use the whole time every week but we'll see. For about 1 to 1.5 hours every I will give a lecture presenting material. Some weeks we will also discuss and article that was assigned to read. Then we will have a lab session for 1 to 1.5 hours where you will work through some problem which will likely be very similar to the problem set you have as homework. In addition we will schedule an open lab one evening early in the week (probably Monday or Tuesday) where Annie will be available to help you.

2.1 Problem Sets (35%):

The only way to learn quantitative analysis is by doing. Thus most weeks there will be a problem set / homework assignment. The problem-set is due Wednesday at noon before class. That will give us time to grade them an return them to you in class. You are encouraged to work on the problem sets in groups but everyone must hand in an own writeup. You should also note with whom you have collaborated on the problem set.

Problem-sets are not graded but you are required to complete them all. If you have a wrong answer on a problem set you will get the chance to re-do/correct that part of the problem set. You will only get full credit if you successfully complete all problem sets (including re-dos) and you will only be able to re-do problem sets that you attempted at the first due date.

2.2 Participation (10%):

This means coming prepared to class and labs and being a good colleague. The learning in this class should be a collaborative endeavor. We will be available for consultation during labs and office hours and also over e-mail. To facilitate the helping each other out you should post any questions you have on the discussion board on the my courses website.

2.3 Replication paper (55%):

The main requirement for this course is a research paper. This process will work in two steps. First you will find a recently published (approx. last 5 years but certainly no more than 10 years) research paper in your field of interest. You will then replicate the analysis in that paper. The replication part is due roughly mid-course. For the final paper your task is to extend upon or improve the paper you replicated. You will do this in teams of 2 (if

we have an uneven number of students we may have one team of 3 or one solo author).

It is imperative that you start working on this NOW.

By January 28 at the latest you should have identified a co-author and by *February 11* you both bust have decided on a research paper that you intend to replicate. Sometime before Feb 11 you *must* consult me about the paper you chose by that date. You don't have to take my advice but you still must consult me. For the consultation make an appointment with me and send me a pdf of the paper you intend to work on.

On *March 17* you will send your replication (including R code) to another team of students who will replicate your replication and provide helpful feedback on it. Your *constructive* feedback will be worth 10% of the course grade.

The last class will be reserved for presentations of your papers (5% of grade). The quality of the presentation will be part of the grade for the research paper. The final product is due one week after the last class (April 21) and will be worth 40% of the final grade.

2.4 Presenting statistical results:

One learning objective of this course is how to professionally present statistical analysis. Again the only way of learning this is practice and developing good habits. As thus I will be **very** strict in enforcing this throughout the course. This includes all homework assignments and the replication paper. Sloppy presentations will not be accepted (see re-do policy above). Generally this will mean journal quality tables, graphs and writeup of your result. We will establish early in the course what constitutes acceptable level of presentation and we will practice this throughout the class. Suffice it to say for now that copy-pasting raw regression output will not pass.

It is not required but I **strongly** recommend that you use this opportunity to learn LaTeX- a free typesetting software that will create professionally formated papers and allow you to focus on what is important - the content. Also LaTeXallows you to typeset mathematical formulae much more straightforwardly than other word processors. As with R there is a learning curve but the payoff is well worth it. Once you got the hang of it you will realize how inferior and clumsy MS Word is for writing academic papers. LaTeX nicely integrates with reference management software (e.g. BibDesk) that

will keep your journal articles filed for you.

We will offer a brief introduction to LaTeX early in the course.

2.5 Key Dates

- January 14: First course meeting
- January 28: Find co-author
- February 11: Settle on paper to replicate.
- March 4: no class - reading week: Replication, replication
- March 17: Replications due
- April 14: Last class - presentation of final projects
- April 21: Final papers due at 4pm. Mode of delivery TBD

3 Topics

We will go through the material as fast as possible provided that everyone in the class can follow. As such the outline below should be understood as a rough marching plan not something that we will follow adhere to at any cost. Some topics will require more than one week to cover and others we may move through more slowly. I will most likely update and add to the list of readings for some weeks as we move along. We will, at minimum, cover regression models for continuous data (OLS) and some models for discrete data (logistic regression). If we have time remaining at the end of the course we can cover additional topics based on student interest (and instructor expertise).

NOTE: I haven't received my copy of the new edition of the Fox textbook yet so some of the chapters are tentative/to be assigned.

1 Introduction

Outline of the course, what is quantitative analysis about? introduction to R and LaTeX

- King, G. (2006). Publication, publication. *PS: Political Science & Politics*, *39*(01), 119–125

- Young, C. (2009). Model uncertainty in sociological research: An application to religion and economic growth. *American Sociological Review*, *74*(3), 380–397
- Broockman, D., Kalla, J., & Aronow, P. (2015). Irregularities in lacour (2014). Tech. rep., Stanford University (Not required but certainly an interesting and disturbing story. You can find more on various internet gossip sites. Also a good illustration for replication analysis)
- Fox 1

2 What is regression analysis?

Conditional expectations, local averaging, functional forms, interpretations

- Tatem, A. J., Guerra, C. A., Atkinson, P. M., & Hay, S. I. (2004). Athletics: momentous sprint at the 2156 olympics? *Nature*, *431*(7008), 525–525
- Fox Chapter 2

3 Bivariate regressions

Properties of OLS, estimation, interpretation...

- Fox 5.1 and 6.1

4 Visualizing and presenting data

Examining data, introduction to lattice package in R, aesthetics and best practices

- Healy, K., & Moody, J. (2014). Data visualization in sociology. *Annual review of sociology*, *40*, 105–128
- Tufte, E. R. (2001). *The visual display of quantitative information*. Cheshire, Conn.: Graphics Press, 2nd ed ed
- Cleveland, W. S. (1993). *Visualizing data*. Murray Hill, N.J.: AT&T Bell Laboratories
- Fox Chapters 3 and 4

5 Multivariate Regression (2 weeks)

Estimation and interpretation, dummy variables and interactions

- Braumoeller, B. F. (2004). Hypothesis testing and multiplicative interaction terms. *International organization*, *58*(04), 807–820
- Fox Chapters 5.2, 6.2., 7
- Aachen Chapters 5 and 6 (Course Website)
- Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, *22*(4), 327–339

6 Discrete regression models: Logistic and ordered logistic regression

Motivation, estimation, quantities of interest

- Mood, C. (2010). Logistic regression:why we cannot do what we think we can do and what we can do about it. *European Sociological Review*, *26*(1), 67–82
- Fox 14

7 The Maximum Likelihood approach to statistical inference

Stochastic and systematic components, likelihood functions, optimization

- King, G. (1998). *Unifying political methodology : the likelihood theory of statistical inference*. Ann Arbor: University of Michigan Press
- Fox 15

8 Discrete regression models continued

Nominal variables, count variables, censored variables and more, presenting and interpreting results

- Fox 14, 15
- King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American journal of political science*, (pp. 347–361)

9 Assessing Model Adequacy

Statistical assumptions, substantive knowledge and model building

- King, G., & Roberts, M. E. (2014). How robust standard errors expose methodological problems they do not fix, and what to do about it. *Political Analysis*, (p. mpu015)
- Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, *22*(4), 327–339
- Fox 11,12
- Western, B. (1991). A comparative study of corporatist development. *American Sociological Review*, *56*(3), 283–294

10 Matching

Model dependency, exact matching, propensity scores...

- TBD

11 Missing data

Multiple imputation

4 Policies

Academic Integrity: McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offenses under the Code of Student Conduct and Disciplinary Procedures.

Submitting Written Work in French: 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.