SOC252H1: Intermediate Quantitative Methods

Department of Sociology

University of Toronto

Fall 2020

***Lecture Day/Time**:* Tuesday 10am – 12pm

Mode of delivery: Online synchronous via BB Collaborate

Course website: Quercus

Tutorials:

Wednesday 4-6pm and Friday 12-2pm

Instructor

Monica Alexander

Email: monica.alexander@utoronto.ca

Office Hours: Tuesday 12:30pm – 2:30pm

Book office hours via: https://calendly.com/monica-alexander/office-hours

***Teaching Assistant*:**

Abtin Parnia

Email: abtin.parnia@mail.utoronto.ca

Course Description

As social scientists, we are interested in understanding how social outcomes vary across different groups, or how such outcomes are related to other characteristics and variables of interest. To answer such questions, we often need to collect data and analyze that data in a statistical way. The course introduces fundamental techniques and methods to analyze quantitative data to draw inferences about

social processes. Specifically, the course covers exploratory data analysis, data visualization, simple and multivariate linear regression, and logistic regression analysis. Issues of estimation, inference and method diagnostics will be discussed. The emphasis of this course is not only to learn how to apply statistical techniques, but also to identify data issues that could potentially bias results.

Prerequisites

The prerequisite to take this course is SOC202H1 (Introduction to Quantitative Methods in Sociology). Students without this requirement will be removed at any time discovered and without notice. In general, students are expected to have a solid background in univariate statistical analysis, including the basics of probability and statistical inference.

Texts

There are no required textbooks for this class. Each week, supplementary readings will be indicated. Many of the topics covered in this course are from

- Gelman, Andrew; Hill, Jennifer, and Vehtari, Aki. 2020. 'Regression and Other Stories'. (listed as GHV in readings)

In addition, some recommended readings are drawn from these references, which are freely available online or available online through the UofT library:

- Grolemund, Garrett and Wickham, Hadley. 2020. 'R for Data Science' (https://r4ds.had.co.nz/). (listed as R4DS in readings)
- Illowsky, Barbara and Dean, Susan. Collaborative Statistics. 2008. Connexions: Online. (https://cnx.org/contents/XgdE-Z55@40.9:LnCgyaMt@17/Preface) (listed as CS in readings)
- Navarro, Danielle 'Learning Statistics with R' (https://learningstatisticswithr.com/lsr-0.6.pdf)

- Ware, William; Ferron, John; and Miller, Barbara. 2013. 'Introductory Statistics, A Conceptual Approach Using R' (available online through library) (listed as WFM in readings)

All other readings will be posted on Quercus.

Mode of Delivery

This course will be held entirely online. Lectures, tutorials and office hours will occur synchronously through BBCollaborate. Given the current circumstances, I understand it may be difficult to attend class for a number of reasons, and as such all lectures and tutorials are being recorded such that they can be accessed at a later time.

Software

All statistical computing for this course will be done using R (https://www.r-project.org/about.html). R is a statistical programming language, and computations are executed from a set of typed commands. The best way to use R is through RStudio, an editor which allows you to better see your code, directory and output: https://rstudio.com/products/rstudio/.

Both R and RStudio are free to download to your own personal computer

1. Download R here: https://cran.rstudio.com/
2. Download RStudio here (free version) https://rstudio.com/products/rstudio/download/

Data

We will be using data from the 2017 General Social Survey for Canada in many parts of this course. This is a long-term survey run every year by Statistics Canada

on a representative sample of Canadians aged 15 and over. We will provide you will a version of the GSS with already constructed variables. We will also make available a codebook for the GSS which lists all available variables in the raw data. If you want to create your own variables from the raw data for the research project, you may do so in consultation with the instructor or TA.

Evaluation

Homework Assignments (45%)

Throughout the semester, you will be asked to complete three homework assignments based on material covered in lecture. Each assignment will have about 5 to 10 questions that will ask you to apply a method using R with real data and interpret the results. You are expected to complete these assignments individually, although some consultation among classmates is normal and expected. Your assignments should be completed in RMarkdown, and the submission should include the knitted pdf and RMarkdown file containing the necessary code to produce the results in your assignment. Assignments will be submitted electronically via Quercus. Each of these assignments will count for 15 percent of your final grade, and thus altogether, they will count for 45 percent of your final grade. Late homework assignments will not be accepted and will receive a mark of zero.

In-class Midterm Test (20%)

The midterm test will be held during class time online on 27 October. It will cover all lectures and tutorials from September 15 to October 20 and will count for 20 percent of your final grade. You will have the full class period (110 minutes) to complete the test, which will consist of both multiple choice and short-answer questions that may involve some calculations.

Research project (35%)

In addition to assignments, each student will develop a research question of their choice using data from the Canadian General Social Survey (GSS), address it by using the descriptive and inferential techniques presented in the course to analyze data in R, and write a short report summarizing your findings.

There will be several milestones throughout the semester to help you smoothly progress towards a final report. I will give you feedback at each step of the way. Your overall grade on the research paper will be the sum of your grades on the following assignments:

- **Research question and choice of dataset (5%):** state the research question, why we might care about the answer to this question, hypotheses about what you think the answer is to the research question (and why), and the key independent and dependent variables in the dataset that you will use to answer this question.
- **Descriptive statistics, exploratory data analysis (EDA), statistical analysis plan (5%):** Discuss with the aid of tables and graphs, the main characteristics of your variables of interest. Show how these descriptive and exploratory data analyses help to inform your statistical analysis plan. The code to produce all results in this part should also be submitted, in either R file or RMarkdown file format.
- **Statistical analysis (5%):** Perform the relevant statistical analysis or analyses on your dataset in order to gain insight into your research question. Include a discussion on the choices you made to come up with your final statistical model, and some discussion and interpretation of your results. The code to produce all results in this part should also be submitted, in either R file or RMarkdown file format.
- **Final report (20%)**: The final write-up of the report, which is worth 20 per cent of your total grade, is due at the end of semester, and brings together all parts you have completed throughout the term in a complete report. It is expected that the report be written in the style of an academic paper,

including the following sections: Abstract, Introduction, Data, Methods, Results, Discussion, and Conclusion. A report template with guidelines on word length for each section will be provided.

In sum, the project counts for a total of 35 per cent of your final grade. Each component should be completed in RMarkdown, and the submission should include the knitted pdf and RMarkdown file containing the necessary code to produce the results. All parts should be handed in electronically via Quercus.

Provisional due dates for required work are as follows:

Date	Assessment due	Weight (%)
6 October	Assignment 1	15
	Project: Research question, choice of variables	5
27 October	Mid-Term	20
3 November	Assignment 2	15
	Project: Exploratory data analysis	5
1 December	Assignment 3	15
	Project: Statistical analysis	5
15 December	Project: Final write-up	20

Course Policies

Attendance

You are expected to attend every class and arrive online in a punctual manner. However, as mentioned above, given the current circumstances, I understand it

may be difficult to attend class for a number of reasons, and as such all lectures and tutorials are being recorded.

Communication

The best way to ask questions about course material or assignments is in person during your instructor's office hours. The following are guidelines for email communication with your TA and the course instructor: please make sure that you have a legitimate need before you write and that you cannot resolve your question during office hours; email messages should state the course number and the purpose of the email clearly in the subject line.

Late Homework Assignments

If you are unable to turn in an assignment or miss the test for medical reasons, you will need to email me the instructor, not the TA, and also declare your absence on ACORN, within one week of the missed assignment/test. For other reasons, such as family or other personal reasons, please contact your college registrar and have them email me directly.

Re-marking

We will use specific marking keys assignments. Those keys define the universe of possible answers and possible variations in those answers. In a course such as this, the only issue that may come up is a mistake in applying the key to the answers in specific cases. If there is a mistake in an assignment or test you get back, you should see the TA within two weeks of your receipt of the assignment. In general, we will not consider work for re-grading after feedback on a later test or assignment, unless it is in this two-week period.

Accessibility

If students require accommodations or have any accessibility concerns, please visit http://studentlife.utoronto.ca/accessibility as soon as possible.

Academic Misconduct

Academic integrity is fundamental to learning and scholarship at the University of Toronto. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that the University of Toronto degree that you earn will be valued as a true indication of your individual academic achievement, and will continue to receive the respect and recognition it deserves. Cheating, misrepresentation, and plagiarism will not be tolerated. Students who commit an academic offence face serious penalties. Know where you stand by reading the "Code of Behaviour on Academic Matters" in the Calendar of the Faculty of Arts and Science.

Tutorials

This course has weekly tutorials on Wednesday 4-6pm and Friday 12-2pm. The tutorials are designed to provide instruction in the application of methods covered in lecture using the statistical software package R. Each will involve working through an applied analysis of data as part of your homework assignments with the guidance of a teaching assistant.

Class Schedule

Date	Lecture	Tutorial	Reference readings	Assessment due
15 September	Preliminaries, review	Installing R/RStudio Introduction to R, RMarkdown	**WFM** Ch 1 **CS** Ch 1.1 to 1.10, 2.6 to 2.9 **Navarro** Ch 5.1to 5.2 **R4DS** Ch 4, 5, 27	
22 September	Exploratory data analysis and data visualization	Data visualization in R (ggplot)	**GHV** Ch 2.1-2.3 **WFM** Ch 2-4 **CS** Ch 2.4 to 2.5 **R4DS** Ch 3	
29 September	Probability, sampling distributions and hypothesis testing	Simulation	**GHV** Ch 4 **CS** Ch **3**.1 to 3.4, 5.1 and 5.2, 6.1 to 6.6, 7.1 to 7.5, 9.1 to 9.4, 9.7 to 9.10 **Navarro** Ch 10	
6 October	Simple Linear Regression I	Linear regression in R	**GHV** Ch 1.1-1.4, 6,7,8.1-3 **WFM** Ch 6-7	Assignment 1 Project: research qn
13 October	Simple Linear Regression II	SLR workflow		
20 October	Multiple Linear Regression I	MLR in R	**GHV** Ch 10.1-10.4, 11.1-11.3, 12.1-12.5 **Navarro** Ch 15	
27 October	MID-TERM			
3 November	Multiple Linear Regression II	Transformations, interactions	**GHV** Ch 10.1-10.4, 11.1-11.3, 12.1-12.5 **Navarro** Ch 15	Assignment 2 Project: EDA
10 November	READING WEEK, no class			
17 November	Binary Outcome models I	Logistic regression in R	**GHV** 13.1-13.3, 13.7, 14.1-14.2	
24 November	Binary Outcome models II	Prediction	**GHV** 13.1-13.3, 13.7	
1 December	Model Selection	Model Selection in R	**GHV** Ch 11.7-11.8	Assignment 3 Project: analysis
8 December	Extensions	Regression Discontinuity	Readings on Quercus	
15 December				Research project due