SOC6707H: Intermediate Data Analysis
Department of Sociology
University of Toronto
Winter 2022

Lecture Day/Time: Mondays 10am-1pm
Location: Room 240
Course website: Quercus

Instructor
Monica Alexander
Email: monica.alexander@utoronto.ca
Office hours: TBC

Teaching Assistant:
Julia Ingenfeld
Email: julia.ingenfeld@mail.utoronto.ca
Office Hours: TBC

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 builds on fundamental techniques and methods to analyze quantitative data to draw inferences about social processes. Specifically, the course covers linear regression and generalized linear regression, and several advanced regression topics. 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

It is assumed that students have successfully taken SOC6302H (Stats for Sociologists) or equivalent. If you are not a sociology student please contact me to get approval to enroll in the course.

Texts

There are no required textbooks for this class. If you are looking for supplementary readings, many of the topics covered in this course are from

- Gelman, Andrew; Hill, Jennifer, and Vehtari, Aki. 2020. 'Regression and Other Stories'

Some resources to help with learning R that are freely available online:

- Grolemund, Garrett and Wickham, Hadley. 2020. 'R for Data Science' (https://r4ds.had.co.nz/).
- Alexander, Rohan. 2020. 'Telling Stories with Data' (https://www.tellingstorieswithdata.com/).

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/

Evaluation

Homework Assignments (30%)

Throughout the semester, you will be asked to complete two 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 30 percent of your final grade. Late homework assignments will be subject to a penalty of 10 percentage points per day.

In-class Midterm Test (20%)

The midterm test will be held during class time online on 26 February. It will cover all lectures and labs from in the first 5 weeks and will count for 20 percent of your final grade. You will have 2 hours to complete the test, which will consist of both multiple choice and short-answer questions that may involve some calculations.

Research project (50%)

In addition to assignments, you will develop a research question of their choice using a dataset of your choice, 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), the dataset that you will be using to answer this question, and the key independent and dependent variables in the dataset that you will use to answer this question. It is expected that you will have downloaded your data and read it into R.

- **Descriptive statistics, exploratory data analysis (EDA), statistical analysis plan (10%):** 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 (10%):** 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.

- **Presentation (5%):** Present short summary of your findings in the final class of semester.

- **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.

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.

Course Policies

Attendance

You are expected to attend every class and arrive in a punctual manner.

Communication

The best way to ask questions about course material or assignments is in person during the lecture, lab or 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 the lecture, lab or 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, 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 me 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 Services

It is the University of Toronto's goal to create a community that is inclusive of all persons and treats all members of the community in an equitable manner. In creating such a community, the University aims to foster a climate of understanding and mutual respect for the dignity and worth of all persons. Please see the University of Toronto Governing Council "Statement of Commitment Regarding Persons with Disabilities" at http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppnov012004.pdf.

In working toward this goal, the University will strive to provide support for, and facilitate the accommodation of individuals with disabilities so that all may share the same level of access to opportunities, participate in the full range of activities that the University offers, and achieve their full potential as members of the University community. We take seriously our obligation to make this course as welcoming and accessible as feasible for students with diverse needs. We also understand that disabilities can change over time and will do our best to accommodate you.

Students seeking support must have an intake interview with a disability advisor to discuss their individual needs. In many instances it is easier to arrange certain accommodations with more advance notice, so we strongly encourage you to act as quickly as possible. To schedule a registration appointment with a disability advisor, please visit Accessibility Services at http://www.studentlife.utoronto.ca/as, call at 416-978-8060, or email at: accessibility.services@utoronto.ca. The office is located at 455 Spadina Avenue, 4th Floor, Suite 400.

Additional student resources for distressed or emergency situations can be located at distressedstudent.utoronto.ca; Health & Wellness Centre, 416-978-8030, http://www.studentlife.utoronto.ca/hwc, or Student Crisis Response, 416-946-

7111.

Academic Integrity Clause

Copying, plagiarizing, falsifying medical certificates, or other forms of academic misconduct will not be tolerated. Any student caught engaging in such activities will be referred to the Dean's office for adjudication. Any student abetting or otherwise assisting in such misconduct will also be subject to academic penalties. Students are expected to cite sources in all written work and presentations. See this link for tips for how to use sources well:

(http://www.writing.utoronto.ca/advice/using-sources/how-not-to-plagiarize).

According to Section B.I.1.(e) of the Code of Behaviour on Academic Matters it is an offence *"to submit, without the knowledge and approval of the instructor to whom it is submitted, any academic work for which credit has previously been obtained or is being sought in another course or program of study in the University or elsewhere."*

By enrolling in this course, you agree to abide by the university's rules regarding academic conduct, as outlined in the Calendar. You are expected to be familiar with the *Code of Behaviour on Academic Matters*

(http://www.artsci.utoronto.ca/osai/The-rules/code/the-code-of-behaviour-on-academic-matters) and *Code of Student Conduct*

*(*http://www.viceprovoststudents.utoronto.ca/publicationsandpolicies/codeofstudentconduct.htm) which spell out your rights, your duties and provide all the details on grading regulations and academic offences at the University of Toronto.

Normally, students will be required to submit their course essays to Turnitin.com for a review of textual similarity and detection of possible plagiarism. In doing so, students will allow their essays to be included as source documents in the Turnitin.com reference database, where they will be used solely for the purpose of detecting plagiarism. The terms that apply to the University's use of the Turnitin.com service are described on the Turnitin.com web site.

Equity and Diversity

The University of Toronto is committed to equity and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect. As a course instructor, I will neither condone nor tolerate behaviour that undermines the dignity or self-esteem of any individual in this course and wish to be alerted to any attempt to create an intimidating or hostile environment. It is our collective responsibility to create a space that is inclusive and welcomes discussion. Discrimination, harassment and hate speech will not be tolerated.

Additional information and reports on Equity and Diversity at the University of Toronto is available at

 http://equity.hrandequity.utoronto.ca.

Labs

This course has weekly labs which will be run in the last hour of the session every Friday. 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	Lab	Assessment due
10 January	Preliminaries	Installing R/RStudio Introduction to R, RMarkdown	
17 January	Probability! Chance! Randomness!	Making a report in RMarkdown	
24 January	Exploratory data analysis and data visualization	Getting data and loading it into R	
31 January	EDA and data viz II	Data visualization in R (ggplot)	Assignment 1 Project: Research Question
7 February	Linear Regression I	Linear Regression in R	
14 February	Linear Regression II	Linear Regression: displaying output	
21 February	READING WEEK, no class		
28 February	In-class mid term		
7 March	Logistic regression	Logistic regression in R	Project: EDA
14 March	Logistic regression	Logistic regression: displaying output	
21 March	Multinomial regression	Multinomial regression	
28 March	Ordinal regression	Ordinal regression	Assignment 2 Project: analysis
4 April	Project presentations		
11 April			Research project due