SOC6707 – Intermediate Data Analysis, Winter 2024
Monica Alexander (monica.alexander@utoronto.ca)

Office 9135 (Level 9, 700 University)
Fridays, 9-12, Room 17146

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, generalized linear regression, and multilevel models. 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.

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'

In addition, a helpful resource (particularly for R) is:

Alexander, Rohan. 2023. '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://posit.co/products/open-source/rstudio/.

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

1. Download R here: https://cran.rstudio.com/
2. 2. Download RStudio here (free version): https://posit.co/downloads/

Course Requirements

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, many of which 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 Quarto or RMarkdown, and the submission should include the knitted pdf and Quarto or 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 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 on 16 February. It will cover all lectures and labs from in the first 6 weeks and will count for 20 percent of your final grade. You will have 2 hours to complete the test, which will consist of short-answer questions that may involve some basic calculations. You will not be required to write R code, but may have to interpret R code / output given.

Research project (35%)

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. Your overall grade on the research paper will be the sum of your grades on the following assignments:

-   **Research question and choice of dataset, and short EDA (10%)**: state the research question, 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. Show one or two key graphs or tables that help to inform your analysis.
-   **Presentation (5%)**: Present short summary of your findings in the final class of semester.
-   **Final report (20%)**: The final write-up of the report, should be structured as a short scientific report, and include and Introduction, description of Data, Description of methods of analysis, Results, and Discussion.

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

Course Policies

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 for 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.

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* (https://governingcouncil.utoronto.ca/secretariat/policies/code-behaviour-academic-matters-july-1-2019) 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 www.ouriginal.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 www.ouriginal.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 www.ouriginal.com service are described on the www.ouriginal.com web site.

Accessiblity 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.

Equity and Diversity Statement

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.

Use of Generative AI in Assignments

In general, students are discouraged from using generative AI tools, such as ChatGPT4 in this course. Specific course policies around the use of generative AI tools are as follows:

- **R code:** Students may wish to use generative AI tools to aid in initial development and writing of R code to answer assignment questions and carry out the research project analysis. If this is the case, the use of such tools should be explicitly acknowledged in the submitted work, and the relevant prompts should be included as an appendix to the assignment. In addition, all code must be thoroughly commented or described in the submitted work. Failure to do so may result in penalties.
- **Written work:** Using generative AI tools to generate written answers to assignment questions of text contained in the final research project is prohibited in this course. Representing as one's own an idea, or expression of an idea, that was AI-generated may be considered an academic offense in this course.

This course policy is designed to promote your learning and intellectual development and to help you reach course learning outcomes.

Course Schedule

Date	Lecture	Lab	Assessment due
12 January	Preliminaries, review of probability and sampling	Making sure everyone is fine with R/RStudio/Quarto Data viz review	
19 January	(Quantitative) Research methods and design	More data viz	
26 January	Review of linear regression	How to do linear regression in R	
2 February	Logistic regression	Logistic regression in R	Assignment 1
9 February	Generalized linear models I	Poisson regression in R	
16 February	MID TERM		
23 February	READING WEEK, no class		
1 March	Generalized linear models II	More GLMs	Assignment 2
8 March	Multilevel Models I	Getting started with brms	
15 March	Multilevel Models II	Presenting results	
22 March	Multilevel regression and poststratification	MRP	Project: RQ and EDA
29 March	GOOD FRIDAY, no class		Assignment 3
2 April	Make up class (TBD)		
5 April	Research presentations and recap		
12 April			Research project due