SOC6302H: Statistics for Sociologists

Fall 2021
Lecture: Mondays, 1-3 pm
Lecture Location: Online for first 2 weeks, then TBD
Lab: Mondays, 3-4 pm
Lab Location: Online for first 2 weeks, then 36

Instructor: Andrew Miles
Office location: 364
Email address: andrew.miles@utoronto.ca
Office hours: by appointment

Teaching Assistant: Julia Ingenfeld, julia.ingenfeld@mail.utoronto.ca

Course Description

This class is designed to help you traverse the entire gamut of data analysis, from conceptual understanding of statistical methods to practical application, so that you can become a savvy consumer of statistical information and be able to start answering your own questions. To that end, lectures and discussions will be interspersed with exercises and assessments designed to aid learning and retention, and students will participate in a lab tutorial each week.

Course Objectives

You will…
- learn the basic concepts and skills needed to begin answering research questions using quantitative data
- apply these concepts and skills to answering research questions
- learn to use statistical software for statistical analyses

Textbooks and Other Materials

The textbook for the course is:

Cumming, Geoff and Robert Calin-Jageman. 2017. *Introduction to the New Statistics: Estimation, Open Science, & Beyond*. New York: Routledge.

Ritchie, Stuart. 2020. *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*. Metropolitan Books.

There are a number of texts out there that teach R, but everything you will need to know will be taught in lab. For additional information (or reference in case you forget something) you can consult the following resources:

For information on the R language and working with data…

Wickham, Hadley, and Garrett Grolemund. *R for Data Science*. Available at: https://r4ds.had.co.nz/.

For introductory statistics texts that use R, try the following. I have not read them, but I have a high level of confidence that they will more or less cover what we do in class, since most introductory statistics classes cover very similar material …

Sonderegger, Derek L., and Robert Buscaglia. 2020. *Introduction to Statistical Methodology, Second Edition*. Available at: https://bookdown.org/dereksonderegger/570/1-summary-statistics-and-graphing.html.

Wehde, Wesley, Hank Jenkins-Smith, Joseph Ripberger, Gary Copeland, Matthew Nowlin, Tyler Hughes, Aaron Fister, and Josie Davis. *Quantitative Research Methods for Political Science, Public Policy and Public Administration for Undergraduates: 1st Edition With Applications in R*. Available at: https://bookdown.org/wwwehde/qrm_textbook_updates/.

Statistical Software

All statistical analyses will be performed using R, which can be freely downloaded at https://cran.r-project.org/. You should also download and install RStudio (here: https://www.rstudio.com/products/rstudio/download/).

Calculator

You will need a scientific calculator or the equivalent to perform calculations. A key pedagogical component of this course is grasping the underlying logic of statistical procedures by working them out by hand.

Class Format

The course has been designed to maximize your ability to learn, retain, and apply statistical concepts by drawing on principles gleaned from research on learning and the brain. In particular, it takes seriously the idea that brains need repetition by providing repeated opportunities to practice statistical skills.

Lecture time will be devoted to presenting statistical concepts. Lectures will be held live over Zoom for the first several weeks of class, and then—if all goes well—live in an actual classroom after that.

You are encouraged to take notes by hand. To help you focus on learning the material rather than scrambling to take notes, simplified versions of lecture slides will be posted prior to lecture. It is recommended that you print a copy of these slides and use them during the lecture to take notes on. We will take a 10-minute break about halfway through each lecture period. Lectures will

typically also include one or more in-class exercises designed to help you solidify key concepts and practice statistical skills. However, the real learning will occur outside of lecture periods as you participate in lab tutorials and complete readings and assignments.

Lab Tutorials

The goal of lab tutorials is to review material from lecture as needed and learn to apply statistical concepts to real data problems using software designed for statistical analysis. Labs are led by the course TA and will typically include an in-lab exercise to help you practice your data analytic skills. Labs will be held live using Zoom for the first few weeks of class and (hopefully) in person thereafter.

Weekly Reading and Assignments

Reading assigned chapters from the textbook and completing assignments provide useful repetition and practice of concepts and skills presented during lecture.

Reading is intended as review, and therefore should occur *after* lecture, but before the following lecture. In this way, lectures can provide a framework that will help students better understand and retain the material they read.

Evaluation and Grading
Coursework is weighted as follows in calculating the final grade.

Weekly assignments	45%
Research briefs	30%
Final Test	25%

Final grades will be assigned using the grading scale below, taken from the University Assessment and Grading Practices Policy.

Percentage	Grade
90-100	A+
85-89	A
80-84	A-
77-79	B+
73-76	B
70-72	B-
0-69	Fail

Description of Evaluation Components

Weekly Assignments

There will be an assignment posted on Quercus each week that will allow you to review key concepts and practice data analysis. Assignments will often include a computing component using R, the how-to's of which will be taught in lab sessions.

Assignments can be found under *Course Materials* → *Assignments*. When you open an assignment for a given week, you will see two things. The first is a PDF file of the assignment, including any instructions, labeled "Assignment # - questions," where # is replaced by the assignment number. The second is a link that allows you to submit your answers. *All answers must be submitted via this link* (unless otherwise noted in the assignment instructions).

Each week, a portion (or all) of the assignment will be graded by Quercus, and so the score will be immediately available on Quercus. Students who are not satisfied with their score are permitted to correct their work and resubmit it (one time only per assignment). Any corrections must be resubmitted by the original due date. Sections of the assignment that are graded by the instructor (not Quercus) may be submitted only once.

Research Briefs

You will complete two research briefs during the course. The goal of each research brief is to give you a chance to apply the skills you've been learning to a question that is of interest to you.

<u>What to Do</u>

1. Find a topic that interests you and come up with some aspect of that topic that you'd like to learn more about. For example, if you are interested in education, you might ask: "How much education do different types of people in Canada get?" Or you might ask, "Do different ethnic groups get different levels of education?" Keep in mind that you will need to be able to answer this question using techniques you've learned in class. This means that your question should be simple.
2. Find some data that will allow you to answer your question. A good place to start is the General Social Survey (Canadian or American). NOTE: If you are having trouble finding data, the easiest thing to do is probably to pick a new question, one that can be answered using the data you have access to.
3. Figure out which variables you will need to answer your question.
4. Determine which statistical technique you can use to answer your question.
5. Run the analysis in R.
6. Interpret the results. That is, what is the answer to your question? How do you know? What evidence do you have?
7. Write up the results in the Research Brief (see guidelines below)
8. Turn the Research Brief in.
9. Celebrate your budding statistical prowess in a manner of your choosing.

<u>Specific Requirements</u>

Research Brief #1 will need to use one or more techniques from weeks 1-5 of the course (see the course schedule later in the syllabus).

Research Brief #2 will need to use one or more techniques from weeks 6-10.

A Research Brief should be no longer than 2 pages, with normal (12 point) sized font. Often a single page will be sufficient.

Include at least one table, graph, or figure to help display your results. This should be formatted and presentable, not simply a screenshot of output from R.

Use the layout described below for your Research Brief (i.e., all the same headings in the same order). The goal is to mimic the essential elements of an academic publication.

Layout

Your Research Brief should include
- **Question:** a clearly worded question
- **Data:** a description of the data you are using, including a description of all variables and how they are coded
- **Plan of Analysis:** a description of your analysis (i.e., how are you going to use the data to answer the question?)
- **Results:** a presentation of the results, including a table/figure/graph
- **Discussion:** Provide an interpretation of the results – that is, given the results, what is the answer to the question? Are there any other possible interpretations? What limitations might there be in your data or analysis that affect how well we can answer the question?

Choosing a Good Research Question: Research questions come in many levels of generality. Consider the following:

Level 3	Is racial inequality a problem?
Level 2	Are there racial disparities in educational outcomes?
Level 1	Do people of Asian descent get less education than Whites?
Analytic level	Does a greater proportion of Whites than Asians obtain a bachelor's degree?

The level 3 question is more like a research program—you can imagine trying to answer this question in lots of different settings, with different racial groups, and so forth. That is, the level 3 question could be the motivation for many different papers. The level 2 question looks at a specific type of racial inequality—disparities in educational outcomes. This is much more focused than level 3, but you can still imagine trying to answer this question in a variety of ways (and in a number of different papers). The level 1 question is a focused, empirically tractable attempt to answer the level 2 question. It is the type of question that could serve as the focus for a single research paper.

Your research brief should use a level 1 question. If it is helpful to you, you may couch your level 1 question in terms of a level 2 question, e.g., "Do Canadians experience racial disparities

in educational outcomes? To begin to answer this, in this research brief I will examine whether Asians have lower educational attainment than Whites." But you are also welcome to simply state your level 1 question.

Be careful not to use an "analytic level" question. An analytic level question is actually a *description* of how you will answer your level 1 question that is sneakily disguised as a question. You will translate your question into specific techniques in the "Plan of Analysis" section.

Examples of research briefs can be found on Quercus.

Links to submit research briefs can be found on Quercus under *Course Materials → Research Briefs*.

Grading

Research Briefs will be graded as follows:

Score	Meaning
3	Excellent
2	Adequate
1	Needs attention

For the purposes of final grade calculation, scores on Research Briefs will be averaged and assigned the following values:

Score range	Grade	Numeric value
0	Fail	0
0-1	Fail	50
1-1.3	B-	71
1.4-1.6	B	75
1.6-1.7	B+	78
1.8-1.9	A-	82
2-2.4	A	87
2.5-3	A+	95

Final Test

The final test is a take-home affair designed to assess all of the skills developed during the course, with a particular focus on apply analytic skills to realistic research questions using real data. Students should therefore be able to
   1) develop an appropriate plan of analysis for a given research question
   2) perform the analysis
   3) interpret the results
The test is open book and open notes, but should be completed individually (i.e., not discussed with other people). The weekly assignments and research briefs are designed to prepare you for the exam.

Course Schedule

All readings are from the textbook unless otherwise noted.

Week	Date	Topic	Reading (C & CJ)	Due this week (by the start of lecture)
1	Sept 13	Data – levels of measurement, variables, populations, samples, data quality	chapter 2	nothing
2	Sept 20	descriptive statistics	chapter 3	assignment 1
3	Sept 27	probability distributions, normal distributions, and sampling distributions	chapter 4	assignment 2
4	Oct 4	Dealing with Uncertainty –confidence intervals	chapter 5	assignment 3
5	Oct 11	Dealing with uncertainty – hypothesis testing; one sample tests	chapters 6-7	assignment 4
6	Oct 18	Dealing with uncertainty – more hypothesis testing; two sample tests	chapters 6-7	assignment 5 research brief #1
7	Oct 25	Analyzing data – measures of association for categorical variables, tables and chi-square	chapter 13	assignment 6
8	Nov 1	Analyzing data – covariance, correlation, and bivariate linear regression	chapters 11-12	assignment 7
9	Nov 15	Multivariate relationships – the logic of controls	Agresti chapter 10	assignment 8
10	Nov 22	Multiple linear regression	Agresti chapter 11	assignment 9
11	Nov 29	Interaction terms	Brambor et al. 2006	assignment 10 research brief #2
12	Dec 6	Open science	chapter 10; *Science Fictions* (read before class)	assignment 11
	Dec 12	Final test (stage 1) due by midnight		
	Dec 16	Final test (stage 2) due by midnight		

Every attempt will be made to follow this schedule, but it is subject to change at the discretion of the instructor.

Procedures and Rules

Rounding

Unless otherwise specified, round all answers to 2 significant figures. This means to report the first two digits that carry any real meaning. For example, all of the following are rounded to two significant figures: 45.23, 0.34, 0.00044.

Late Work

Late work is not accepted. However, an individual student's due date for an assignment can be extended for a legitimate reason with proper documentation (e.g., illness, family emergency, religious observance, but NOT family vacations, weddings, I want a long weekend, etc.). All requests for extensions are at the discretion of the instructor. Where possible, these arrangements must be made in advance of the missed work. If that is not possible, the student should contact the instructor as soon as is feasible. Work submitted before an extended due date is not considered late and incurs no penalties.

Grade Appeals

If you believe that a mistake was made in grading your work, you may appeal the grade by submitting a written explanation of why you think your mark should be altered to the instructor. The instructor will then re-grade your work with the additional information in mind. Although in most cases re-grading results in a higher mark, this is not guaranteed, and your mark might go down.

Academic Integrity

You are expected to abide by the University's standards of academic integrity, which can be found in the "Code of Behaviour on Academic Matters". Plagiarism or other violations will be addressed in accordance with University guidelines. Please be cautious in this matter, as violations of academic integrity must be referred to the Dean's office and the penalties can be quite severe.

Working with Other Students

Working with other students is often a useful way to learn statistics. You are therefore encouraged (but not required) to work with other class members in completing assignments (including research briefs). However, each student must complete and submit his/her own work, written in his/her own words. Students who work together on class work should also indicate whom they worked with on each assignment (if anyone). These steps guard against situations where a student's academic integrity might be called into question (see section on Academic Integrity).

Attendance

Data analysis is a skill, and like any skill mastering requires time on task. Attendance is therefore mandatory at all lectures and lab tutorials. Any absences should be cleared with the instructor. More than two absences at either lecture or lab may result in a reduction in a student's final grade, usually one-half grade per absence beyond the two (e.g., from A to A-). These grade

reductions will be at the instructor's discretion. Students who cannot view lectures and labs live should consult with the instructor so that special accommodations can be made.

Copyright Notice

Course materials prepared by the instructor are considered by the University to be an instructor's intellectual property covered by the *Copyright Act*, RSC 1985, c C-42. These materials are made available to you for your own study purposes, and cannot be shared outside of the class or "published" in any way. Lectures, whether in person or online, cannot be recorded without the instructor's permission. Posting course materials or any recordings you may make to other websites without the express permission of the instructor will constitute copyright infringement.

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, especially when discussing topics that might evoke strenuous disagreement or other strong feelings. Class members need not agree with each other or the instructor but must refrain from speech or other behavior that creates an intimidating or hostile environment. Additional information and reports on Equity and Diversity at the University of Toronto is available at http://equity.hrandequity.utoronto.ca.

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". In working toward this goal, the University will strive to accommodate 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.

Last updated 7-7-21