POL222H1F: Introduction to Quantitative Reasoning I

University of Toronto, St. George

Summer 2022

Instructor: **Catherine Moez**, PhD Candidate, Department of Political Science

Contact: **catherine.moez@mail.utoronto.ca**

Office Hours: TBD

Class Format: In-person. Class sessions will be recorded for those who cannot attend.

Meeting Room: SS 2118

Meeting Time: Tuesdays and Thursdays 6:00pm-8:00pm (May 10th to June 16th)

Teaching Assistants: TBD

Tutorial Sessions: TBD (Fridays 4pm, 5pm, 6pm)

Course Description and Objectives

Many things we interact with in our daily lives are automated, from software that predicts the next words you will use in a text message to plagiarism-detection programs, the algorithms that shape your social media feeds, and more. Quantitative research approaches, including both observational and experimental designs, are also common in academic political science and policy studies. This course is an introduction to the statistics and regression models that underpin these approaches. You will learn about key foundational concepts in statistics, such as extrapolating from a data sample to a full population of cases, making causal claims, testing relationships between variables, and summarizing data.

The course will cover:

- Foundations of statistical inference (population vs. sample; signal vs. noise in data; the sampling distribution, and other general concepts in significance testing). Relationship of data analysis to causal theory testing.
- Univariate statistical measures (mean; median; mode; variance; standard deviation; frequency distributions).
- Bivariate statistical measures (correlation; covariance; etc.). Bivariate significance tests.
- Observational and experimental research designs.

- Ordinary least squares (OLS) linear regression, at an introductory level. Definitional overview of linear regression; uses of regression for theory testing and prediction; assumptions and implications of assumption violations. Interpretation of regression tables and the regression line equation.
- While not central to this course, you will also see previews of other regression models, such as those with categorical dependent variables (logistic regression; preview of machine learning models). We will also look at data visualizations throughout.

This knowledge underpins other methodology courses you may wish to take in your studies, such as POL232H1 Introduction to Quantitative Reasoning II.

Textbook

There is no textbook for the course. Required or recommended readings will be posted on Quercus, under the Readings Module.

Software

No software is required for assignments and tests. If any math calculations are required, it will be possible to perform them with a basic calculator (default phone or computer calculator; Excel; R; etc.). I may show examples of data analysis in the free software program R (r-project.org) in class, but you do not need to download or use it for this course. (POL232H1 will require you to use a program such as SPSS, Stata, or R to conduct your own data analysis).

Assignments

- **Week 1 Assignment**: Variables, Univariate Measures, Introductory Concepts.
  Covers material from lectures 1 and 2; due at start of lecture 3 (May 17th). (**15%** of mark).
- **Week 2 Assignment**: Bivariate Relationships, Hypothesis Testing.
  Covers material from lectures 3 and 4; due at start of lecture 5 (May 24th). (**15%**).
- **Midterm quiz**: Covers material from lectures 1-6. Causal theory; univariate and bivariate statistics; significance testing. (May 31st, written online during class time). (**15%**).
- **Study Review**: Review of a policy institute paper and/or an experiment-based study. (To be assigned to you). Focus on causal claims; sample representativeness; concept operationalization; (if applicable) random assignment to treatment and treatment design. Review strengths and weaknesses; propose improvements. (June 7th). (**15%**).
- **Political Science Article Review**: Review of a regression table and the causal arguments advanced in an academic political science article. (To be assigned to you). (June 14th). (**15%**).
- **End of term test**: Research designs; Linear regression; all course material. (June 16th). (**15%**).
- **Attendance and participation** (**10%**). This mark is entirely for tutorial attendance and participation, but up to 2 percentage points on your final grade can be allocated as a bonus for in-lecture participation (polls/question answers during lecture time). (+2%).

There is no final exam.

We will have consistent policies throughout all tutorials for absences, late penalties, and extensions. Please see the end of the syllabus for more details.

Topic Schedule

Week 1: Foundational Concepts and Univariate Statistics

(Variables, Relationships, Univariate Statistics)

Lecture 1 (Tuesday, May 10th). Introduction to the course. Variables (Definitional). Types of Variable/Levels of Measurement. Concept Validity. Concept Operationalization.
(Lecture Part 2) Introduction to Causation in Social Science. "IV" and "DV", Causal Theories and Hypotheses.

> Qs: Levels of measurement; Hypothesis wording; Causal theory vs data measurement (Intro).
>
> Required Reading: KKV (King, Keohane and Verba 1994, *Designing Social Inquiry*). Chapter One: "The Science in Social Science."

Lecture 2 (Thursday, May 12th). Univariate Statistics. Central tendency measures (mean, median, mode). Dispersion measures (variance, standard deviation, IQR). Frequency distributions. (In real data, and ideal types). Math notation for the univariate measures.
(Lecture Part 2) Population vs. Sample. Descriptive Statistics vs. Causal Inference (Theory testing and describing relationships in data, cont.). 'Populations' in social science. Sample collection – the sampling frame, random sampling, other sample collection techniques.

> Qs: Types of variable—connection to the measures. 'Population' meanings in social science. Math notation for univariate measures.
>
> Required Reading: KKV Chapter 3, "Causality and Causal Inference."

Week 2: Bivariate Statistics

(Bivariate Statistics and Introduction to Significance Testing)

**Week 1 Assignment due at the start of Tuesday's class (May 17th, 6pm), on Quercus or on paper.*

Lecture 3 (Tuesday, May 17th). Bivariate descriptive statistics: Correlation, covariance, other association measures. (Formulas and math notation, cont.). Bivariate X-Y plots. (Different linear and nonlinear relationships; non-relationships). Dataset view vs. plot view. Math notation for bivariate measures.
(Lecture Part 2): Some informal tests and initial steps in determining whether a relationship exists between two variables. Contingency tables. Difference in group means. Time permitting: Concept of signal vs. random noise in statistics; Probability Basics.

Qs: Measure difference in group means (absolute difference, without the significance test). Measure proportion of each group with a specific DV outcome from a contingency table. Informally, does this seem to be noise or signal to you?

Readings: TBD. (Possibly: Kellstedt and Whitten chapter on causality).

**Week 1 Assignment due at the start of Tuesday's class (May 17th, 6pm), on Quercus or on paper.*

Lecture 4 (Thursday, May 19th). General introduction to statistical significance testing. Sample vs. population and concept of noise, reviewed. The sampling distribution. Basic steps in conducting a significance test. Interpretation (results; 'confidence level').
(Lecture Part 2): Specific example of a bivariate significance test: Correlation coefficient significance test; difference in means t-test; chi-squared test. Interpretation (p-value, critical t, alpha/confidence level). Connecting results of a hypothesis test to causal theory.

Qs: When would we use each test? (Types of variable review).

Readings: TBD.

Week 3: Significance testing, continued

(Bivariate Significance Testing, Multiple X Variables (introduction), Review for Midterm Quiz)

**Week 2 assignment due at the start of Tuesday's lecture, 6pm May 24th. Submit via Quercus or on paper.*

Lecture 5 (Tuesday, May 24th). Bivariate significance testing cont.: Difference in means t-test; Chi-squared test.

Qs: Establishing the null hypothesis in different examples.

Readings: TBD (Possibly: Johnston 1970 on the Student's t-test).

**Week 2 assignment due at the start of Tuesday's lecture, 6pm May 24th. Submit via Quercus or on paper.*

Lecture 6 (Thursday, May 26th). Multiple X variables; correlation matrix and relationships between IVs. Midterm Quiz review. Math concepts and math notation review (Percentages and proportions; percent change; order of operations; indexing; mathematical transformations). (Time permitting: History of statistics).

Qs: Review: Correlation vs. causation. Problematic causal claims. IV and DV too conceptually close? Concept of correlated IVs.

Readings: TBD (Possibly: Wooldridge on controlling for variables).

Week 4: Research Designs

(Observational, Experimental, and Natural Experiment/Quasi-Experiment research designs)

**Midterm Quiz on all material from Weeks 1-3 (Lectures 1 through 6) will be held online at the start of Tuesday's class. (6:10-6:45pm).*

Lecture 7 (Tuesday, May 31st). Observational research designs. Definition. Common types: opinion surveys, other observed data (economic data, etc.). Time-series, cross-sectional, and time-series-cross-sectional (panel) designs.
(Time permitting: Contemporary problems with sample collection; Recent polling failures; Weighting and other strategies to correct for nonrepresentative samples).

Qs: The causality problem. Omitted variables? Regression to the mean? Correlation alone vs. an actual causal mechanism? Time trends?

**Midterm Quiz on all material from Weeks 1-3 (Lectures 1 through 6) will be held online, on Quercus, at the start of Tuesday's class. (6:10-6:45pm).*

Lecture 8 (Thursday, June 2nd). Experiments. General concept of experiments: Randomized assignment; researcher-controlled treatment; comparison of outcomes. The "Fundamental problem of causal inference". Internal and external validity, potential problems with treatment design.
(Lecture Part 2): Types of experiment: Randomized controlled trials (Lab experiments); Survey experiments; Field experiments (Wantchekon); Natural experiments (AKA quasi-experiments). Pros and cons of observational research in comparison to experiments.

Qs: Are natural experiments experiments? (Job training programs, Policy changes, ...).

Readings, observational research: Nevitte, Neil, 1996, "Setting the Stage." In *The Decline of Deference*, Broadview Press.

Readings, experiments: TBD (Possibly: Kahnemann and Tversky; Campbell 1969; Huntington-Klein).

Week 5: Ordinary Least Squares (OLS) Linear Regression

(Bivariate and multiple linear regression)

**Study Review due at the start of Tuesday's class, 6pm June 7th. (Via Quercus).*

Lecture 9 (Tuesday, June 7th). Bivariate linear regression. Basic concept of regression with two variables. "Ordinary least squares" meaning; line-fitting to data. Reading line equations. Visualizing regression lines. Interpreting regression tables – slope and intercept.
(Lecture Part 2): Connection between regression estimation from sample data, and theory testing. Reminder that potential causal claim problems still apply – regression math is essentially correlational. Uses of regression for theory-testing and prediction.

Qs: Filling in predicted values using the line equation. Visually recognizing different types of relationship.

**Study Review due at the start of class (6pm, June 7th) on Quercus.*

Lecture 10 (Thursday, June 9th). Multiple regression. Multiple X variables in linear regression: General concept of 'controlling for' or 'adjusting for' other X variables. Interpretation of slopes when multiple X variables are in the regression model. Explained variance; R-squared.
(Lecture Part 2): Significance of regression coefficients. Uncertainty of estimates. (Sample vs. population). Time permitting: Regression assumptions and impact of assumption violations. Omitted variables and impacts of excluding or including relevant X variables.

Qs: Interpretation of dummy variables. Reading tables from academic publications.

Readings: TBD.

Week 6: Linear Regression cont.; Other Regression Models

(Linear regression assumptions; Logistic regression; Preview of other quantitative approaches)

**Political Science Article Review due at the start of Tuesday's class, 6pm June 14th. Submit on Quercus.*

**End of term test will be held, online, at the start of Thursday's class, 6:10-6:45 pm June 16th.*

Lecture 11 (Tuesday, June 14th). Linear regression, continued. Interaction terms. Assumption violations, and implications of assumption violations, in more detail. (Bias, consistency, and efficiency; other vocabulary you may see used around regression: 'Estimator', 'Regress Y on X', etc.). R-squared and Adjusted R-squared as measures of 'goodness of fit'. Time permitting: Some strategies for nonlinearity. (Logged X variables, Squared X variables).
(Lecture Part 2): Other regression models: Logistic regression. Main differences: Y interpretation; Slope interpretation.

> Qs: Impact of assumption violations. Interpreting the DV in logistic regression. Interaction terms. Discussion and review of second part of the course.
>
> Readings: TBD.
>
> **Political Science Article Review due at the start of Tuesday's class.*

Lecture 12 (Thursday, June 16th). Preview of other quantitative approaches. Logistic regression extensions, such as: Multinomial logistic regression. Preview of machine learning approaches (Basic concepts; Uses; Automatic feature selection). Time permitting: Unconventional data (image, audio, text); Data visualizations.

> Readings: Breiman, Two Worlds.
>
> **End of term test will be held, on Quercus, at the start of class (6:10-6:45pm).*

Administrative Policies

Important Dates

For an up-to-date listing of the Faculty of Arts and Science (FAS)'s important dates for Summer 2022, please check https://www.artsci.utoronto.ca/current/dates-deadlines/academic-dates#academic-dates-deadlines-accordion-6. As of April 1st, 2022, the deadline to withdraw from a summer F term course is May 31st, 2022. Per FAS guidelines, you will receive at least one piece of marked work, worth at least 10% of your grade, by May 31st, 2022. There is no final exam, so our class meetings will end with the June 16th session (final test and lecture).

Assignment Formats

Quizzes/tests will be written on Quercus, during class time. Assignments will also be submitted on Quercus. For each assignment, please submit one written document, in Word (.doc, .docx) or PDF format. The first two assignments will require some math calculations, and so you can either: (a) Write the math by hand and take photos/screenshots of it, to include in your Word document; (b) submit these first two assignments in paper format in class; or (c) write out the math equations using a program such as LaTeX or R Markdown (online version: https://upmath.me/), then copy images of your equations into your document. Any of these options are acceptable, as long as the content is clear and readable. If you do include photos of hand-written math in your submission, the photo must be clear (lighting, size of the text, legible handwriting, etc.).

Plagiarism and Academic Integrity

Students are expected to complete their own work, to be familiar with University of Toronto plagiarism/academic integrity policies, and to abide by them. Failure to be aware of a plagiarism-related rule is not considered acceptable grounds for breaking it.

In this course specifically, consultation with other students is permitted in terms of understanding and discussing the material, but you must write up your own original assignments. (This means that you may discuss the interpretation of regression coefficients or similar technical questions with other students and your TA, but you cannot submit the same or highly similar written text as other students for a written assignment.) You must explain your answers in your own original words. For math questions, you must show the steps of your calculations.

Quizzes are written online, meaning that consulting your class notes or Internet sources is possible and acceptable. However, the quiz is expected to reflect your own work. Do not share your answers with other students. (Quiz questions may also have slight variations or shuffled response order, to prevent copying).

The Plagiarism Detection Tool:

The two lengthier written assignments, the Study Review and the Political Science Article Review, will be submitted to the University's plagiarism detection tool. TAs and the instructor also reserve the right to manually check for similarities between different students' work, for all assignments. Please keep copies of your notes and draft work for these two written assignments until the end of the course (i.e., the time of final grades being submitted to the department), in the event that you are requested to show your work process.

University statement on the plagiarism detection tool:

"Normally, students will be required to submit their course essays to the University's plagiarism detection tool 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 tool's reference database, where they will be used solely for the purpose of detecting plagiarism. The terms that apply to the University's use of this tool are described on the Centre for Teaching Support & Innovation web site (https://uoft.me/pdt-faq)." (https://teaching.utoronto.ca/ed-tech/teaching-technology/pdt/pdt-faq/).

Please note that students are allowed to opt out of using the plagiarism detection tool, without being penalized for this decision. Alternative checks on the academic integrity of the submitted work, such as showing draft material or answering questions about the research process, may be required in this circumstance. Please notify the instructor if you wish to opt out. (Source: https://teaching.utoronto.ca/ed-tech/teaching-technology/pdt/pdt-faq/).

General Absence, Missed Test and Late Assignment Policies

We will have consistent policies throughout all tutorials for absences, missed quizzes, late assignment submissions, and extension requests.

For in-person tutorial absences, this will involve completing the same set of participation questions as those asked of students in the in-person tutorial, for each session missed. For the bonus participation mark relating to participation in lectures, a similar set of questions will also be available on request for any class session missed, which will count the same as in-person participation.

Quizzes/tests are written online (through Quercus) during a limited window of class time. The test must be written during this time. If you must miss a test for a legitimate reason (announced either ahead of time or after the fact – as early as is feasible), you will be asked to write a similar but alternative test at another time.

Late assignments will receive a penalty of **3 percentage points per day**. If you will submit your assignment after the point when on-time assignments have been returned to the class (generally 10-14 days after the deadline), *you will be asked to complete a similar but alternative assignment instead*. **Submitting the original assignment after others have been marked and returned to students** (with feedback) **will not be accepted (grade of 0)**, so please contact me if you think you will be submitting an assignment past the 10 day mark/return date, so that you can receive the correct alternative assignment as early as possible. **This policy applies to all work submitted after the time when the original assignment has been returned to the class, whether the late submission has an extension or not**.

For assignments, please contact me or your TA as early as possible if you will need an extension (waived late marks) for a legitimate reason. Legitimate reasons include illness; family issues; last-minute work or childcare reschedules; and other unforeseeable and/or unavoidable issues. Extensions will not be granted for late submissions due to circumstances that are within your control and foreseeable, such as failing to plan your coursework schedule between multiple classes. (Please also back up copies of your files, and if you have an Internet or computer problem

close to a deadline, please try to document the situation and submit your work as soon as possible once the issue is resolved). We will generally not require documentation of your situation for personal and family issues.

If you have an ongoing accommodation with Accessibility Services (https://studentlife.utoronto.ca/department/accessibility-services/) that will affect your ability to meet standard assignment deadlines throughout the term, please let your TA know as early as possible before a deadline you may potentially miss. Similarly, if you require accommodations on quizzes/tests for disability-related reasons, please inform me (the instructor), or have a representative from Accessibility Services inform me on your behalf. We may require documentation from Accessibility Services in this instance (i.e., if you are requesting changes to quiz/test format, or requesting an ongoing Accessibility Services extension on assignments throughout the term).

Accessibility (Other Modifications)

Separately from assessments of your work, please inform me if there are other modifications you will need in order to access and understand course material, e.g., issues with colour-coding, lecture audio, visuals/images in lecture slides, note-taking in class, etc.

In-Person Attendance Policies

Please note that that course quizzes/tests will be conducted online during the time of the class, meaning that you do not need to physically stay in the classroom to write these assessments. Other assignments will be submitted via Quercus as well. Recordings of the in-class lectures will be posted on Quercus as soon as possible after class. If at any point you feel that you should not be physically present in class or tutorials for Covid-related reasons (personally sick or suspect you may be sick; recovering; seriously concerned about your health or household members' health), **please do not attend in person. You will not be penalized for missing participation/attendance for this reason.** For any in-person tutorials missed, you will be asked to answer a few brief questions for participation marks, the same as those asked of students in the in-person tutorial. (Marked largely for completion/effort, not correctness). There are also alternative questions you can complete for the in-lecture participation bonus grade of up to 2%, if you miss any in-person lectures.

Other Grading Policies: Reassessment

If you believe that any of your work has been marked incorrectly, and are requesting a re-grade, you will need to write out your reasoning and justification for a higher grade in a formal email to your TA. Please take some time to consider the feedback first, i.e., wait at least one day, but also send the request soon after the assignment has been returned, i.e., within about 2 weeks). The request must be based *entirely* on the substance of the work you submitted, in relation to the assignment requirements, and not on other factors such as effort or your usual performance. Your TA will then re-assess the work. Please note that your mark may stay the same, go up, or go down after reassessment. If there is any further disagreement, I will then re-assess the submission.