Statistical Analysis for International Affairs

Norman Paterson School of International Affairs, Carleton University

Syllabus version: December 6, 2019

Instructor:	Dr. Oskar Timo Thoms	Course number:	INAF 5016
Email:	timo.thoms@carleton.ca	Term:	Fall 2019
Office Hour:	Monday 10:00-11:30	Class Time:	Thursday 11:35-14:25
Office:	Richcraft Hall 5118	Class Location:	Tory Building 360

Contents

1 Course Description

Assessing many questions in international affairs requires the ability to think analytically about quantitative data and statistics. This course offers an introduction to quantitative analysis and research methods in political science, international relations, and conflict analysis, and will cover causal inference, probability theory, and statistical estimation. The focus will be on hands-on data analysis and practical application of basic statistical methods. The prerequisite is MA standing in the Norman Paterson School of International Affairs or permission of the School.

2 Learning Goals and Teaching Method

This course is on applied quantitative methods, not on mathematical statistics. While students will perform calculations using computer software, the emphasis is on the comprehension and practice of basic quantitative methods. The goals of this course are for students to:

- learn the basic concepts and tools of descriptive and inferential statistics;
- be able to intelligently and critically consume statistics in academic and popular literature;
- implement and interpret standard statistical methods in their own research;
- and acquire a foundation for possible further study of applied methods.

Class time consists of 13 weekly lectures and tutorials led by teaching assistants (TAs). Lectures will last 2 hours and 50 minutes, including a short break. Each student will be assigned to one weekly tutorial session, and must remain in that one for the semester. Instead of exams, there will be weekly assignments and regular problem sets, in which students analyze and report on datasets, thereby integrating knowledge of class concepts, practical analysis, and communication of results.

Students will learn and use the statistical software package R for applied course work, and will be introduced to spreadsheet software, such as Microsoft Excel or Googlesheets, as a data management tool used in conjunction with R.

R is open-source, free, and more importantly, is at the cutting edge of applied statistics and data manipulation and visualization in academia and industry. R has a steep learning curve because it is a programming language and there are usually multiple possible solutions to a given programming task. However, R is also a good learning tool for practical data analysis because it brings the user close to the data and is great for interactive data analysis. This course will take a minimalist approach in order to flatten the R learning curve. Students will learn the basics required to do the course work. By focusing on the most important features and essential functions for conducting analyses, this will provide enough knowledge of R to complete a research project and a foundation for further study of more advanced use if desired.

3 Course Requirements and Evaluation

The final course grade will be made up of:

- weekly assignments for practicing R programming and illustrating concepts to be submitted during Weeks 3-12 (totaling 10%)
  - due on Tuesdays at noon
- participation in weekly tutorials (10%);
  - all tutorials are held on Thursdays
- 3 problem sets applying concepts and conducting data analysis (15% each, totaling 45%)
  - due Fridays of Week 5 (Oct 4), Week 8 (Nov 1), and Week 11 (Nov 22), at noon
- a final project to be submitted as a take-home exam (35%)
  - due Friday, Dec 13 at noon

Further details on assignments, problem sets, and the final project are provided in Section 7.

4 Required Reading and Software

The course will use the following textbook, and follow its teaching philosophy:

Kosuke Imai. 2018. *Quantitative Social Science: An Introduction*. Princeton University Press.

The paperback book is available from the Carleton University Bookstore, or from www.amazon.ca and www.chapters.indigo.ca. Electronic access can be rented or purchased from VitalSource, Perusall, and others. The first two chapters (needed for Weeks 1-3) are available to download for free (in PDF) on the publisher's website.

It is strongly recommended that students install the R and RStudio Desktop software packages on their own laptops and bring these to tutorials. The first tutorial (Week 1) will cover software setup. Any students who are not able to bring a laptop to tutorials, should contact their assigned TA as soon as possible, and during the first tutorial at the latest.

5 cuLearn and Communication

This course uses cuLearn, Carleton's learning management system. To access your course on cuLearn go to http://carleton.ca/culearn. For help and support, go to http://carleton.ca/culearnsupport/students. Any unresolved questions can be directed to Computing and Communication Services (CCS) by phone at 613-520-3700 or via email at ccs_service_desk@carleton.ca.

Assignments, problem sets and the final project will be submitted through cuLearn, and grades will be posted on cuLearn. Instructor **announcements** may be made in class or on cuLearn. This syllabus is a living document; any changes will be announced and the most recent version will always be available on cuLearn. If you seek clarification on course logistics, please post your question in the *open Q&A* forum under the *Syllabus and Course Logistics* section in cuLearn, so that any responses are available to all students. While I will provide clarification when appropriate, **anyone** in the course is allowed and encouraged to answer questions in this forum.

There is separate **discussion forum** under the *Discussions* section in cuLearn. Here students are encouraged to ask each other questions about substantive course concepts and R programming and respond to their peers' questions.

The best way to talk to me outside of class time about anything related to the course is during the weekly office hour. If you want to talk in person but are unable to make the office hour due to a legitimate scheduling conflict, please send me an email, suggesting at least two alternative times, to schedule an appointment. Please note that I live outside of Ottawa and my access to campus on non-teaching days will be very limited.

Student must use their Carleton email account for all course-related correspondence, and put "INAF5016: " at the beginning of the subject line for email filtering. If you do not do this, your email may go missing during busy times. I check my email frequently during the week, but less so on weekends, and I do not guarantee an instant response. As a general rule, I will try to respond within 1-2 business days (not on weekends). Please do not wait until the last minute to email me with questions, problems, or concerns, particularly before course deadlines; I simply will not be able to respond to a flood of emails the night before an assignment is due or a couple of hours before class.

Please feel free to give me feedback on the course and my teaching. **Really**. I will be open to constructive comments and criticism.

6 Class Expectations

6.1 Advance preparation

Students are expected to complete the assigned readings for each week before class so that they are able to complete the review exercises (using *swirl*[1]) and assignments, and participate in tutorials.

Note that learning statistics and methods cannot be "crammed the night before," and doing so effectively depends on active engagement. It is crucial that students work through the materials continuously and cumulatively, as each week we will build on previous weeks' materials. Start problem sets early in case they prove challenging. Do not fall behind working through the materials. If you start to do so, discuss this with your teaching assistant immediately, in order to plan a course of action for catching up.

6.2 Attendance

Lecture attendance is required for learning the course content. While this attendance cannot be monitored and enforced in such a large course, I will **not** respond to queries outside of class about course logistics which have been comprehensively covered in class.

Tutorial attendance is mandatory and will be monitored and included in the participation grade. Unexcused absences and tardiness count against the participation grade since it is not possible to participate when absent. **Students with three or more unexcused absences from tutorials**

[1] *Swirl* is a useful platform for teaching R programming and data science interactively within the R console. We will explain in lecture or tutorials how to use it.

automatically fail the course. Consult section 8 for the policy on documented excuses. If you have to miss a tutorial, it is important that you consult your TA at least 48 hours in advance.

6.3 Classroom courtesy

In addition to being respectful toward all in the class, please observe the following courtesies:

- Arrive **on time** and do not leave early.
- Turn **off or mute** your cell phone before the beginning of class.[2]
- Do not engage in side conversations.
- Use technology such as laptops, tablets and phones **only** for class purposes.

6.4 Active participation

Learning is an active and collaborative process; we learn through dialogue and collective discussions. Tutorials will be devoted to reviewing and elaborating concepts from lectures and the readings, and applied data analysis tasks in R. Students are strongly encouraged to ask questions about the materials and help answer the questions of their peers, both in tutorials and on the discussion forum. Good participation does not require students to have the correct answers, but to constructively engage with the course materials and show evidence of active learning.

6.5 Student response system during lectures

At times, we will use the Poll Everywhere online platform for student involvement during lectures. Students will send their responses in real time from their smartphones, laptops or tablets, using text messages (SMS), a website, or a dedicated phone app. This will be introduced and explained during the first lecture. Those students who want to install the app on their phones should visit www.polleverywhere.com/mobile.

The website for all in-class polls will be PollEv.com/THOMS. Note that the polls are set up such that all responses are **anonymous** and cannot be linked to individual students.

7 Assignments

This section details the nature and requirements of the assignments, and how to electronically submit them through cuLearn. Email or paper submissions will **not** be accepted. Also consult the collaboration policy for all assignments in section 9.4.

During the first weeks of the course, **further details will be added below as needed**.

[2] Cell phone ringing during class is very disruptive; please make sure this does not occur.

7.1 Weekly assignments

During **Weeks 3-12**, students will submit assignments practicing R programming and illustrating concepts. Assignment instructions and questions will be announced on Fridays of the previous week, and the assignments will be due on Tuesdays at noon.

7.2 Problem sets

There will be three longer problem sets applying concepts and conducting data analysis. These will be due on Fridays of **Week 5 (Oct 4)**, **Week 8 (Nov 1)**, and **Week 11 (Nov 22)**, at noon. Problem set instructions and questions will be announced at least one week prior to the due dates.

7.3 Final project assignment

The final project will be submitted as a take-home final exam, due Friday, Dec 13 at noon. This will be either a replication of an existing study or a student-driven research project, roughly equivalent in length to a problem set. More details will be provided later in the term.

7.4 Formatting requirements

During Week 2, I will provide a **template R Markdown file** on cuLearn that students shall adapt for all assignment submissions. In such a file, students will combine their R code, software output and responses, in order to create fully integrated reproducible research documents for submissions. Further details are in the **submissions_template.Rmd** file.

8 Late submissions and documented excuses

Late submissions of weekly assignments and problem sets will **not** be accepted, graded, or counted, except in case of emergencies.

It is the students' responsibility to plan and manage their time (including time to deal with unforeseen technical problems) so that each and all of the assignments will be completed on time to avoid an incomplete grade. Exceptions will be granted only for documented emergencies, such as serious illness or extraordinary circumstances.

Unless university policy explicitly mandates otherwise, excused absences from class or extensions of submission deadlines will only be granted for unforeseeable and urgent circumstances beyond your control that prevent you from fulfilling course requirements on time. In such cases, we will require supporting documentation, such as from a medical professional or the equivalent, of the condition and severity so as to prohibit work. Any excuse based on a mitigating circumstance known to the student before a deadline or session in question, where the student could have informed us, **must** be approved in advance.

Extensions are possible for the final project, but requests must be in writing and arrangements must be made with your TA **before** the due date. Late submissions of the final project (without an extension) will be penalized one letter grade increment per day or part thereof.

9 Academic Integrity

9.1 Plagiarism

The University Senate defines plagiarism as "presenting, whether intentional or not, the ideas, expression of ideas or work of others as one's own." This can include:

- reproducing or paraphrasing portions of someone else's published or unpublished material, regardless of the source, and presenting these as one's own without proper citation or reference to the original source;
- submitting a take-home examination, essay, laboratory report or other assignment written, in whole or in part, by someone else;
- using ideas or direct, verbatim quotations, or paraphrased material, concepts, or ideas without appropriate acknowledgment in any academic assignment;
- using another's data or research findings;
- failing to acknowledge sources through the use of proper citations when using another's works and/or failing to use quotation marks;
- handing in substantially the same piece of work for academic credit more than once without prior written permission of the course instructor in which the submission occurs.

Plagiarism is a serious offence which cannot be resolved directly with the course instructor. The Associate Deans of the Faculty conduct a rigorous investigation, including an interview with the student, when an instructor suspects a piece of work has been plagiarized. Penalties are not trivial. They include a mark of zero for the plagiarized work or a final grade of "F" for the course. For further information, visit http://carleton.ca/registrar/academic-integrity/ and download the Academic Integrity Policy (PDF).

9.2 Complementarity

Students are encouraged to build up expertise in areas that may cross multiple courses. It is acceptable to write assignments on related topics. However, you may not simply cut and paste your work from one assignment to another, or essentially submit the same work for two or more assignments in the same or different courses. If you plan on writing on related topics in different courses, you must inform the instructors and discuss what will be acceptable in terms of overlap, and what is not. Failure to notify the faculty members will be viewed unfavourably should there be a suspicion of misconduct.

9.3 Intellectual property

Any materials created for this course by the instructor, teaching assistants, or students remain the intellectual property of the author(s). They are intended for personal use and may not be reproduced or redistributed without the prior written consent of the author(s).

9.4 Collaboration Policy

Programming is an individual creative process similar to writing composition.[3] You must reach your own understanding of a problem and discover a path to its solution. Discussions with other people are permitted and encouraged. However, when it comes to writing the code for your assignment submissions, it must be your own work. Do not, under any circumstances, copy another person's code. Incorporating someone else's code into your program in any form is a violation of academic integrity. Abetting plagiarism or unauthorized collaboration by "sharing" code is also prohibited. Sharing code in digital form is an especially egregious violation; do not e-mail your code or make your source files available to anyone.

Novices sometimes have the misconception that copying and mechanically transforming a program – by rearranging code, renaming variables, or similar operations – makes it something different. Identifying plagiarized source code is easier than one might think. Not only does it quickly identify itself as part of the grading process, but detecting similar code can also be automated with software.

If you have questions about how to use some feature of R, you can certainly ask friends or the teaching assistants, but specific questions about code for assignment submissions must be treated carefully. For each assignment, you must specifically describe in your R source file, whatever help (if any) that you received from others and tell us the names of any individuals with whom you collaborated. This includes help from friends, classmates, and course staff members.

10 Academic Accommodation

You may need special arrangements to meet your academic obligations during the term. For an accommodation request, the processes are as follows:

10.1 Pregnancy obligation

Please contact the instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details, visit the Equity Services website or download the Student Guide to Academic Accommodation (PDF).

10.2 Religious obligation

Please contact the instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details, visit the Equity Services website or download the Student Guide to Academic Accommodation (PDF).

[3] The language in this section is adapted from Marc Ratkovic's quantitative analysis syllabus.

10.3 Academic Accommodations for Students with Disabilities

If you have a documented disability requiring academic accommodations in this course, please contact the Paul Menton Centre for Students with Disabilities (PMC) at 613-520-6608 or pmc@carleton.ca for a formal evaluation, or contact your PMC coordinator to send the instructor your Letter of Accommodation at the beginning of the term, and no later than two weeks before the first in-class scheduled test or exam requiring accommodation (if applicable). After requesting accommodation from PMC, meet with the instructor as soon as possible to ensure accommodation arrangements are made. For more details, visit the Paul Menton Centre website.

10.4 Survivors of Sexual Violence

As a community, Carleton University is committed to maintaining a positive learning, working and living environment where sexual violence will not be tolerated, and is survivors are supported through academic accommodations as per Carleton's Sexual Violence Policy. For more information about the services available at the university and to obtain information about sexual violence and/or support, visit: carleton.ca/sexual-violence-support.

10.5 Accommodation for Student Activities

Carleton University recognizes the substantial benefits, both to the individual student and for the university, that result from a student participating in activities beyond the classroom experience. Reasonable accommodation must be provided to students who compete or perform at the national or international level. Please contact your instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details, see the policy.

You can visit the Equity Services website to view the policies and to obtain more detailed information on academic accommodation at carleton.ca/equity/.

11 Weekly Schedule and Topics

The course is divided into six main sections: Introduction; Causality; Measurement; Prediction; Probability; and Uncertainty. This order differs from some other statistics courses; it is intended to get students started on practical data analysis tasks quickly and cover probability and estimation of statistical uncertainty during the second half of the course when it becomes clearer why these concepts are needed.

The following list is offered as a tentative schedule, and the instructor reserves the right to make changes to topics and learning materials as the semester progresses. The topics for each section are stated under the first week of a given section. Any possible changes to the schedule during the semester would be announced during lecture and reflected in an updated syllabus posted on cuLearn.

11.1 Week 1 (Sep 5): Introduction: Research Design

Topics

- introduction to course
- introduction to research design
- dependent and independent variables
- criteria of arguments and analyses
- differences between qualitative and quantitative methods

Readings

- QSS 1.1-1.2

Recommended Readings (optional)

- John Gerring. 2011. *Social science methodology: A unified framework*. New York, NY: Cambridge University Press.
- Gary Goertz and James Mahoney. 2012. *A tale of two cultures: Qualitative and quantitative research in the social sciences*. Princeton, NJ: Princeton University Press.

11.2 Week 2 (Sep 12): Introduction: Statistics & Causality

Introduction & Causality Topics

- introduction to statistics
- causal effects and the counterfactual (potential outcomes)
- randomized controlled trials (experiments) and the role of randomization
- observational studies and confounding bias
- before-and-after and difference-in-differences designs
- descriptive statistics for a single variable
- central tendency and spread, quantiles
- normal approximation and standard deviation

Reading

- QSS 1.3 & QSS 2
- swirl: INTRO1 & INTRO2 (QSS 1.3)

11.3 Week 3 (Sep 19): Causality (continued)

Topics continued from previous week

Reading

- review QSS 2
- swirl: CAUSALITY1 (QSS 2.1-2.4) & CAUSALITY2 (QSS 2.5–2.6)

11.4 Week 4 (Sep 26): Causality (continued) & Measurement

Measurement Topics

- handling missing data
- summarizing and visualizing the univariate distribution with plots
- survey sampling
- summarizing bivariate relationships
- correlation (and relationship to causation)

Reading

- QSS 3
- swirl: MEASUREMENT1 (3.1–3.4)

11.5 Week 5 (Oct 3): Measurement (continued)

Topics continued from previous week

Reading

- review QSS 3
- swirl: MEASUREMENT2 (3.5-3.7)

11.6 Week 6 (Oct 10): Prediction

Prediction Topics

- scatter plots
- simple linear regression
- least squares
- regression toward the mean

- multivariate linear regression
- model fit

Reading

- QSS 4
- swirl: PREDICTION1 (4.1)

11.7 Week 7 (Oct 17): Prediction (continued)

Topics continued from previous week

Reading

- review QSS 4
- swirl: PREDICTION2 (4.2) & PREDICTION3 (4.3)

Fall Break (Oct 21-25): no class/tutorials

Students are advised to be on track with the course materials by the Fall break. The remainder of the course will be more theoretical and conceptually challenging than the first weeks.

11.8 Week 8 (Oct 31): Probability

Probability Topics

- definition and axioms
- permutations and combinations
- sampling with and without replacement
- conditional probability
- random variables and probability distributions
- expectation and variance
- large sample theorems (Law of Large Numbers, Central Limit Theorem)

Reading

- QSS 6
- swirl: PROBABILITY1 (QSS 6.1–6.3)

11.9 Week 9 (Nov 7): Probability (continued)

Topics continued from previous week

Reading

- review QSS 6
- swirl: PROBABILITY2 (6.4–6.5)

11.10 Week 10 (Nov 14): Probability (continued) & Uncertainty

Uncertainty Topics

- unbiasedness and consistency
- standard error
- confidence intervals
- margin of error and sample size calculation in polls
- analysis of randomized controlled trials
- analysis based on student's t-distribution
- hypothesis testing and statistical significance
- one-sample versus two-sample tests
- power analysis
- regression with uncertainty and inference

Reading

- review QSS 7
- swirl: UNCERTAINTY1 (QSS 7.1)

11.11 Week 11 (Nov 21): Uncertainty (continued)

Topics continued from previous week

Reading

- review QSS 7
- swirl: UNCERTAINTY2 (QSS 7.2)

11.12 Week 12 (Nov 28): Uncertainty (continued)

Topics continued from previous week

Reading

- review QSS 7
- swirl: UNCERTAINTY3 (QSS 7.3)

11.13 Week 13 (Dec 5): Uncertainty (continued) & Modeling

Topics

- more on applied regression modeling
- endogeneity
- unit heterogeneity and fixed effects regression
- beyond linear regression: non-continuous dependent variables
- problems encountered in final projects

Reading

- QSS 8