

## **POL2504H1F: Statistics for Political Scientists**

University of Toronto

Fall 2016

Meeting Room: SS 561  
Lecture: Thursday, 10:00am – 12:00pm  
Review Session: Thursday, 9:20am – 10:00am

Instructor: Kenichi Ariga  
Email: kenichi.ariga@utoronto.ca  
Office: SS 3047  
Office Hours: Tuesday, 10:00am-12:00pm

Teaching Assistant: Anthony Sealey  
anthony.sealey@mail.utoronto.ca

### **Course Description**

This course is a Ph.D.-level introductory course on the theory and application of statistical methods in empirical research in political science. Being the first introductory course on statistics for Ph.D. students, the primary objective of this course is to provide these students with opportunities to acquire the foundational knowledge of statistics needed for further sophisticated statistical methods taught in more advanced courses and eventually self taught in the future.

Today quantitative empirical research is one of the major, standard modes of inquiry in political science. A wide variety of quantitative methods — both in research designs and statistical models — is used throughout the discipline and a significant number of papers using quantitative empirical analysis have been published across subfields.

One important feature of these methods is that they have been growing very fast in their variety and applications. Any single course or even a two-year sequence of classes cannot cover the quantitative empirical methods in its entirety used across the discipline. In fact, the methods of choice would vary depending on your substantive research questions and the type of data you may collect to empirically test your theory. Any serious researcher will be in a position in which s/he should consider which research designs and statistical models would be appropriate to answer his/her substantive research questions. Moreover, a researcher will also be in a position in which s/he should be willing to learn — and occasionally develop — new designs and models to properly answer his/her research questions.

To prepare future political scientists for such situations, an important goal of graduate methods training like this course is *to help students become capable* of considering the appropriateness of various methods in their specific research context and of learning, when needed, new methods both independently and collaboratively with their colleagues. Toward this goal, this class emphasizes acquiring *foundational, theoretical knowledge of probability and statistical inference*, which serves as groundwork for learning advanced quantitative methods and applying the methods appropriately and creatively in the future.

Coverage of the class includes: descriptive statistics, probability theory, statistical inference, and linear regression model.

## **Textbook**

Paul M. Kellstedt and Guy D. Whitten, *The Fundamentals of Political Science Research, Second Edition* (Cambridge University Press, 2013).

### *Recommended Textbook*

Alan Agresti and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences, 4th Edition*. Pearson.

Sean Gailmard. 2014. *Statistical Modeling and Inference for Social Science*. Cambridge University Press.

Jeffrey M. Wooldridge. 2013. *Introductory Econometrics: A Modern Approach, 5th Edition*. South-Western Cengage Learning

## **Computer Software**

Quantitative social science research requires the use of computer and computer software. In this class, you will learn to use a software package called R, which is free to download at <http://www.r-project.org> and is getting popular among many social scientists. By the end of the semester, you are expected to be able to conduct a basic quantitative empirical analysis using R.

Other software popular among political scientists — and social scientists in general — are:

STATA (<http://www.stata.com/>), and

SPSS (<http://www-01.ibm.com/software/analytics/spss/>),

both of which are commercially available. While we will use R in this class, it is advisable to get some familiarity with other software during your doctoral training. There will surely be times during the course of your career when you need to learn how to use a new software package. That is the nature of the work of a researcher and one should always be ready to learn.

## **Class Structure**

This class is designed with a mutually learning community in mind and two-hour weekly lectures are only part of the entire learning experience. Every student is expected to take initiative in his or her own learning. All class requirements are designed to facilitate and help his/her initiative. Students are expected to ask questions proactively during the instructor's lectures and are encouraged to discuss the class materials with each other outside the classroom as well.

There will be weekly homework assignments that are due at the beginning (10:10am) of the next class. Self-study R tutorials will also be assigned in several weeks, mostly at the beginning of the semester. Students are expected to complete these R tutorials before the next class.

There will also be a review session between 9:20am and 10:00am before the lecture, in which you may ask questions on the weekly homework assignments, R tutorials, and any other class materials. Answers to homework assignments will be posted on the class Blackboard site at 9:00am of the class day. You are expected to check the answers before the review session and then ask questions if any during the review session. To save the lecture time for new materials for the week, all questions regarding the homework problems, R tutorials, and the past class materials will be addressed in the review sessions.

### **Blackboard / Learning Portal**

The class Blackboard site (<https://portal.utoronto.ca/webapps/portal/frameset.jsp>) or the Learning Portal will be the primary means through which class announcements and assignments will be distributed. Readings other than the above textbook, lecture slides, and assignments will be made available in the class Blackboard site as well. Its Discussion Board will be the primary method by which you will ask questions about the course materials and get them answered (more on this below). It will be your responsibility to obtain access to the class Blackboard site and regularly check it. There will be an important update to the class Blackboard site at least once a week.

### **Discussion Board**

We will use the Discussion Board in the class Blackboard site as the main medium through which you can ask questions regarding class materials and get answers. Given the nature of the course materials, someone else may have the same question as yours and s/he would benefit from your posting the question and getting an answer through the Discussion Board. You are also encouraged to post an answer to the questions posted by your classmates on the Discussion Board so that we can maintain a mutually-supporting learning community from which all of you benefit. The instructor will regularly check the Discussion Board on Mondays, Wednesdays, and Fridays, and answer questions which have not been adequately addressed by peers. For more complex questions or those that would require an extensive treatment, you are best advised to visit office hours.

### **Collaboration**

Collaboration on a team of multiple scholars is not unusual in contemporary social science research in general, and quantitative empirical political science research in particular. As an introductory course on the methods of such research, this class will provide you with an opportunity to practice scholarly collaboration.

For all assignments, you are encouraged to work in a group. In particular, you are encouraged to submit your work jointly as a group of up to three individuals. Everyone in the group will receive the same mark.

Collaboration may seem burdensome sometimes as it involves time-consuming coordination with other students that you would otherwise not have to go through. However, since collaboration is encouraged and highly valued in this class, your effort to collaborate with your classmates in the data analysis essay and final paper assignments will count toward your participation mark (more on this below).

### **Grading and Evaluation**

Grades in the course will be based on the following items:

1. Weekly Homework Assignments: 15%  
There will be weekly homework assignments throughout the semester (with a few exceptional weeks). The homework questions will assess your understanding of important concepts and methods covered in the class. Some of them may include data analysis using computer software. Each homework assignment will be weighted equally.

All homework assignments will be graded on a pass/fail basis. If the instructor can determine that you gave it reasonable efforts to answer all the questions, you will be given a full credit for that homework, regardless of the number of correct answers. If you do not show a reasonable amount of work, however, your homework will be given a fail or a marginal pass. You will receive no credit in the former case and will receive half the credit in the latter.

All homework assignments are due at the beginning of the lecture (10:10am on Thursdays), and will be returned, preferably during the same lecture, after the instructor recorded their marks (pass, marginally-pass, or fail). Note that the instructor will not mark each individual problem but give a quick scan to determine whether you have shown reasonable effort to answer all questions. You are expected to check your answers based on the answer key posted at the class Blackboard site at 9:00am of the class day and ask questions during the weekly review session in 9:20am-10:00am on the same day before the lecture.

2. R Tutorials: 5%  
There will be a series of self-study R tutorials posted on the class Blackboard site. When a tutorial is assigned, you are expected to complete it before the next class. More specifically, you should follow all the steps specified in the tutorial and understand how they work. You won't need to submit anything but will be asked to self-report whether you have completed each tutorial. Completion of each R tutorial will be weighted equally.
3. Two Data Analysis Essays: 30%  
There will be two essay assignments involving data analysis. In each assignment, you will conduct data analysis using R and write up a short essay based on it. Each essay counts 15% toward your final mark.
4. Final Test: 15%, on the last day of the class (Dec. 1).  
There will be a closed-book final test on the last day of the class. While it is closed-book, you will be given a list of formulas needed for the test.
5. Final Paper: 25%, due Dec. 12 (Mon.).  
You will conduct a linear regression analysis using a dataset of your choice and write a research paper, which addresses the causal theory of your interest based on the linear regression analysis.
6. Class Participation and Collaboration: 10%  
Your class participation mark will be determined by: your participation in lectures and class discussions; your contribution to responding to your classmates' questions on the Discussion Board; and your collaboration (group submission) of the data analysis essays and final paper.

### **Late Penalties and Extension**

All work is late if submitted after the date and time specified as the due date.

- Data analysis essays and a final paper handed in late will result in a penalty of 5-percentage-points reduction per day (e.g., from 90% to 85%). Submitting an essay

within 24 hours from the due date and time will be considered one day late; submitting after 24 hours but before 48 hours will be two days late, and so forth.

- Weekly homework assignments not completed before the class in which they are due will receive a zero grade.

Extension may be made if there is a legitimate reason, such as a medical emergency. You may be requested to provide official documentation that justifies the reason.

### **Office Hours**

You are welcome to visit during the instructor's office hours, which will be held during the time and date specified at the beginning of the syllabus, if you have any questions on the class materials.

There will also be tutorials and office hours held by a teaching assistant. The schedule of TA's tutorials and office hours will be announced later on the class Blackboard site.

### **Email Policy**

If you have questions of personal nature (e.g., accessibility, deadline extension), you may email the instructor and expect a response within two working days. Please start the subject heading of your email with "POL2504:..." I will not answer, however, any questions over email that are of substantive nature concerning the class materials. You will need to post those questions on the Discussion Board, come to the weekly review session before the lecture, or visit office hours to get them answered.

Please note that I will not be able to answer email or Discussion Board questions during weekends.

In the case of your questions of substantive nature on the Discussion Board or those of personal nature over email not answered within two working days (excluding weekends), send me an email to let me know they have not been addressed. Please include "POL2504: Unanswered Question" in the subject heading of your email.

### **Academic Integrity**

You are expected to be familiar with the Code of Behaviour on Academic Matters, available at <http://www.artsci.utoronto.ca/osai/students>, which is the rule book for academic behaviour at the U of T. Potential offenses include, but are not limited to, plagiarism, cheating on tests and exams, and fraudulent medical documentation. All suspected cases of academic dishonesty will be investigated following the procedures outlined in the Code. The consequences for academic misconduct can be severe.

### **Class Topics and Schedule**

The following topics will be covered in the order that they are listed. Corresponding chapters of the textbook are also specified. Please note that the predetermined schedule is bound to be adjusted in this type of course. I will try to follow the schedule below, but be prepared to accommodate adjustments during the semester.

## Week 1 (Sep. 15)

### Class Overview

- Sean Gailmard. 2014. *Statistical Modeling and Inference for Social Science*. Cambridge University Press. Chapter 3 “Observable Data and Data-Generating Process.”

## Week 2 (Sep. 22)

### Part 1. Scientific Study of Politics and Causal Theory

### Part 2. Descriptive Statistics for Univariate Distribution

- Part 1
  - Kellstedt and Whitten, Chapters 1 and 2 (skip 2.6.1- 2.6.2 [but read the introductory portion of 2.6], 2.7, and 2.8).
  - Gary King, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press. Chapter 1 “The Science in Social Science.”
- Part 2
  - Kellstedt and Whitten, Chapters 5.8-5.12

## Week 3 (Sep. 29)

### Part 1. Inherent Difficulties in Empirically Identifying Causal Relationship

### Part 2. Simple Linear Regression As Descriptive Tool

- Part 1
  - Kellstedt and Whitten, Chapter 3
- Part 2
  - Kellstedt and Whitten, Chapters 8.1-8.2

## Week 4 (Oct. 6)

### Part 1. Empirical Research Design

### Part 2. Multiple Linear Regression As Descriptive Tool

- Part 1
  - Kellstedt and Whitten, Chapter 4
- Part 2
  - Kellstedt and Whitten, Chapters 9.1-9.4

## Week 5. (Oct. 13)

### More on Linear Regression As Descriptive Tool

#### Readings:

- Kellstedt and Whitten, Chapters 7.4.3, 8.3, 8.4.1-8.4.3, 9.5
- *Recommended:* Kellstedt and Whitten, Chapters 10.4

## Week 6. (Oct. 20)

### Probability Distribution and Population Model

- Kellstedt and Whitten, Chapter 6.1-6.2
- Jeffrey M. Wooldridge. 2013. *Introductory Econometrics: A Modern Approach, 5th Edition*. South-Western Cengage Learning. Appendix B. “Fundamentals of Probability.”

Data Analysis Essay 1, due Oct.21 (Fri.)

## Week 7. (Oct. 27)

### Sampling Distribution and Point Estimation

- Kellstedt and Whitten, Chapter 6.3 and 8.2

## Week 8. (Nov. 3)

### Interval Estimation

- Thomas H. Wonnacott and Ronald J. Wonnacott. 1990. *Introductory Statistics, 5th Edition*. Chapter 8 “Confidence Intervals.”
- Kellstedt and Whitten, Chapter 6.3-6.5 and 8.4 (introductory paragraph and 8.4.4-8.4.5 only)

**Weeks 9 & 10. (Nov. 10 & 17)**

**Hypothesis Testing**

- Alan Agresti and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences, 4th Edition*. Pearson. Chapter 6 “Statistical Inference: Significance Tests.”
- Janet B. Johnson, H.T. Reynolds, and Jason D. Mycoff. 2015. *Political Science Research Methods, 8th Edition*. CQ Press. Chapter 12 “Statistical Inference.”
- Kellstedt and Whitten, Chapter 8.4.6 - 8.4.8, 8.5, 9.6-9.8

Data Analysis Essay 2, due Nov.11 (Fri.)

**Week 11. (Nov. 24)**

**More on Linear Regression**

- Readings TBA

**Week 12. (Dec. 1)**

**Final Test**

Final Paper, due Dec. 12 (Mon.)