POL2504H1F: Statistics for Political Scientists

University of Toronto
Fall 2014

Meeting Room:	SS 561
Meeting Time:	Friday, 10:00am – 12:00pm

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

Additional Office Hours by Dr. Hanil Chang

Email:	hanil.chang@utoronto.ca
Office:	SS 3045
Office Hours:	Monday, 3:00PM – 5:00PM Thursday, 1:00PM – 3:00PM

Course Description

This course is a graduate-level introductory course on the theory and application of statistical methods in empirical research in political science. It primarily targets Ph.D. students who wish to use statistical methods in their future research. 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 may 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, descriptive, associational, and causal inference, and linear regression model.

Textbook

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

Computer Software

Quantitative social science research requires the use of computers. In this class, you will 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 on your own.

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 problem sets and reading assignments that are due at the beginning of the next class. Answers to a problem set will be posted on the class Blackboard site at the beginning of the class. Before handing in your problem set, you will be asked to check your problem set answers by yourself. We will then devote the first 10-15 minutes of the class time to addressing your questions on the problem set questions.

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.

Grading and Evaluation

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

1. Problem Sets: 40%

   There will be weekly problem sets throughout the semester (with a few exceptional weeks). The problem set questions will assess your understanding of important concepts and methods covered in the class. Some of them will include data analysis using computer software. Each problem set will be weighted equally.

   Collaboration with your classmates, in terms of discussing the problems, is highly encouraged as it facilitates your learning, but "collaboration" by way of dividing up problem set questions is prohibited.

All problem sets will be graded on a pass/fail basis. If you give it reasonable efforts to answer all questions in a problem set, you will be given a pass and a full credit for that problem set, regardless of the number of correct answers. If you do not show a reasonable amount of work, however, your problem set will be given a fail or a marginally-pass. You will receive no credit in the former case and will receive half a credit in the latter.

All problem sets are due in class. You will hand in your problem set after we review the problem set answers at the beginning of the class.

2. Short Essay: 20%, due Mon., Dec. 1.
   You will conduct a linear regression analysis using the dataset of your choice (or you may use the dataset I will make available) and write a short essay, which addresses the causal theory of your interest based on the regression analysis.

   For this essay assignment, you will be allowed and encouraged to work and submit the essay in a group of up to three individuals. Collaboration in a team of multiple scholars is not unusual for contemporary social science research in general, and quantitative empirical political science research in particular. As an essay assignment for a course on the methods of such research, this assignment provides you with an opportunity to practice scholarly collaboration. If you submit your work as a group, everyone in the group will receive the same mark. Please note that both group and individual submissions will be treated equally.

3. Final Exam: 30%, on the final day of the class (Fri, Nov. 28).

4. Class Participation: 10%
   Your class participation grade will be determined by how well you participate in and contribute to the class learning — in particular, 1) asking questions proactively during lectures, and 2) responding to your classmates' questions on the class Discussion Board.

Late Penalties and Extension

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

- Essay assignments handed in late will result in a penalty of 20-percentage-points reduction per day (e.g., from 90% to 70%). 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. Essays handed in more than five calendar days late will receive a zero grade.

- Problem sets 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 an unforeseeable medical emergency. You may be requested to provide documentation.

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.

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.

As I teach four classes this semester, including a large class for undergraduates, I expect many students visiting my office hours. In the past, there were several occasions on which there was a long waiting line of students for my office hours and some students had to leave before they saw me as time had run out. To avoid such a disappointing event, I will maintain a sign-up sheet for my office hours online. Please sign up for a 10-minute block on this sign-up sheet. More details about the sign-up sheet will be posted on the class Blackboard site.

There will also be additional office hours, held by Dr. Hanil Chang, a postdoctoral fellow at the Department of Political Science, during the time and date specified at the beginning of the syllabus. He is also available for appointments upon email request (hanil.chang@utoronto.ca).

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 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, fraudulent medical documentation and improper collaboration on marked work. 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

The following topics will be covered in the order that they are listed. Corresponding chapters of the textbook are also specified. The exact dates when these topics will be covered are not specified on purpose, since the predetermined schedule is bound to be adjusted in this type of course. In principle, I will try to spend a week on each set of topics listed below (e.g., the first week on 1. Introduction, the second week on 2. Descriptive Statistics I, and so on). Be prepared, however, to accommodate adjustments during the semester.

1. Introduction

Readings:

- Gailmard 2014, Chapters 1 and 3.

Subjects:

(1) What Will We Learn and Why?
  - Why do we want to conduct quantitative analysis?
    - Small-N and Large-N Studies
  - Why do we need probability?
    - Probabilistic/Stochastic Relationship
    - Sampling, Theoretical, and Fundamental Uncertainties
  - What is statistical inference and why do wee need it?
    - Sample and Population
    - Data Generating Process (DGP)

(2) Syllabus Review

(3) Introduction to R

2. Descriptive Statistics I: Univariate Distribution

Readings:

- Gailmard 2014, Chapters 2.1-2.2
  - Skip 2.1.2-2.1.3

Subjects:

(1) Classifications of Variable

(2) Description of Univariate Distribution
  - Measurement of Central Tendency: Mean, Median, Mode (MSE, MAE)
  - Measurement of Variability: Variance, Standard Deviation, IQR
  - Rank Statistics (Percentiles, Quartiles)
  - Visualization 1: Histogram, Density
  - Visualization 2: Boxplot

(3) Use R to Describe Univarite Distribution

3. Descriptive Statistics II: Bivariate and Multivariate Distribution

Readings:

- Gailmard 2014, Chapters 2.3-2.4

Subjects:

(1) Crosstabs or Contingency Tables (Joint, Marginal, and Conditional Distributions)

(2) Scatterplots, Covariance, Correlation, and Conditional Mean

(3) Bivariate Linear Regression
  - Derivation and Interpretation

- Linear Probability Model
- Goodness of Fit

(4) Multivariate Linear Regression
(5) Various Specifications of Linear Regression Model (optional: only if time allows)
- Log Transformation, Dummy Variable, Interaction Term

(6) Use R to Examine Bivariate and Multivariate Distribution

4. Probability and Probability Distribution

Readings:

- Gailmard 2014, Chapters 4.1-4.4, and 5.1
  - Skip 4.2.1, 5.1.3, and 5.1.4.

Subjects:

(1) Why Do We Need Probability? --- Model of DGP
(2) Basic Probability Theory
a) Probability, Joint Probability, Conditional Probability
b) Mutually Exclusiveness
c) Independence
d) Bayes' Theorem (optional: only if time allows)
(3) Math Review
- Functions
- Differentiation (Derivatives)
- Integration

(4) Random Variable
(5) Probability Distribution of Discrete Random Variable
- Probability Mass Function
- Cumulative Distribution Function

(6) Probability Distribution of Continuous Random Variable
- Probability Density Function
- Cumulative Distribution Function

(7) Expected Value and Moments
- Expectation or Mean
- Variance and Standard Deviation

5. Probability Distribution of Multiple Random Variables

Readings:

- Gailmard 2014, Chapters 4.5-4.7, 5.2-5.3.

Subjects:

(1) Probability Distribution of Two Random Variables
- Joint, Marginal, and Conditional Probability Distribution
- Independence of Random Variables

(2) Probability Distribution of Multiple ($\geq$ 3) Random Variables
(3) Covariance and Correlation
(4) Conditional Mean and Variance

6. Probability Distribution As Model of DGP

Readings:

- Gailmard 2014, Chapter 6

- Skim 6.2-6.5, and 6.9.
- Skip 6.7-6.8.
- Read the rest as usual.

Subjects:

(1) Probability Distribution As Model of DGP
(2) Conditional Expectation Function (CEF) in DGP
(3) Liner Regression Model
(4) Nonlinear Regression Models --- Parametric Families of Probability Distributions Commonly Used As Models of DGP
(5) Probability Distribution of Observed Data

7. Sampling Distribution and Point Estimation

Readings:

➢ Gailmard 2014, Chapters 7 and 9.2.
- Skip 7.2.4, 7.4.2, 7.5.4, 7.6.2, 7.7, and 9.2.5.

Subjects:

(1) Random Sampling
- Independent and Identical Distribution
- Probability Distribution of Observed Data

(2) Sampling Distribution
- Sampling Distribution of Sample Means (Expectation, Standard Error)
- Central Limit Theorem
- Exact Distribution under Normality of DGP

(3) Criteria on Evaluating Estimators
- Unbiasedness
- Efficiency
- Mean Squared Error
- Consistency

(4) Sampling Distribution of Linear Regression Coefficient Estimates
- Exogeneity of Covariates
- Omitted Variable Bias

8. Hypothesis Testing

Readings:

➢ Gailmard 2014, Chapter 8
- Skip 8.2.6, 8.3.4 - 8.3.6, 8.4, and 8.7.

Subjects:

(1) Standard Hypothesis Tests
- Hypothesis Tests on Linear Regression Coefficients
- Hypothesis Tests on Population Mean
- Hypothesis Tests on Population Proportion
- Asymptotic (Large-N) and Finite Sample (Small-N) Tests
- Comparing Linear Regression Coefficients

(2) Statistical and Substantive Significance
(3) How to Report the Linear Regression Estimation Results
(4) F Tests in Linear Regression
(5) Nonstandard Standard Errors

(6) Multiple Tests and Publication Bias

9. Interval Estimation

Readings:

- Gailmard 2014, Chapter 9.1

Subjects:

(1) Confidence Interval
  - Confidence Interval for Linear Regression Coefficients
  - Confidence Interval for Population Mean
  - Confidence Interval for Population Proportion
  - Asymptotic (Large-N) and Finite Sample (Small-N) Confidence Interval

(2) Hypothesis Testing and Confidence Interval

10. Causal Inference (optional: only if time allows)

Readings:

- Gailmard 2014, Chapter 10

Subjects:

(1) Potential Outcome Models

(2) Selection-on-Observables
  - Matching
  - Liner Regression

(3) Selection-on-Unobservables
  - Instrumental Variable
  - Difference-in-Difference
  - Regression Discontinuity

11. Various Methods of Estimation (optional: only if time allows)

(1) Maximum Likelihood Estimation
- Gailmard 2014, Chapter 9.3 and 9.5.

(2) Bayesian Estimation
- Gailmard 2014, Chapter 9.4.