Introduction to Statistics for Political Science

Oxford University
Department of Politics and International Relations
Michaelmas 2023

Instructors:	Rachel Bernhard | rachel.bernhard@politics.ox.ac.uk
	Leonardo Carella | leonardo.carella@nuffield.ox.ac.uk
GTAs:	Jiani Yan | jiani.yan@sociology.ox.ac.uk
	Scott Singer | scott.singer@politics.ox.ac.uk
	Maksim Zubok | maksim.zubok@nuffield.ox.ac.uk

Lectures:	DPIR Lecture Theatre
	Week 1 and 8 | Wednesday 2.00-4.00pm
	Weeks 2-7 | Wednesday 4.00-6.00pm
Labs:	Lab 1a | Scott Singer | Skills Lab | Friday 9.00-10.30am
	Lab 1b | Scott Singer | Skills Lab | Friday 10.45am-12.15pm
	Lab 2a | Jiani Yan | Seminar Room D | Friday 9-10.30am
	Lab 2b | Jiani Yan | Seminar Room D | Friday 10.45am-12.15pm
	Lab 3a | Maksim Zubok | Seminar Room E | Friday 9-10.30am
	Lab 3b | Maksim Zubok | Seminar Room E | Friday 10.45am-12.15pm

Overview of Course:
This course introduces students to the fundamentals of statistical analysis for the social sciences. We will cover the basics, starting from how we can use statistics to summarize information and describe general patterns of interest to how we can implement predictions or support causal claims. The aim of the course is to make students competent readers and consumers of basic statistical material and competent analysts of quantitative data. In addition, the course will introduce students to code in R to manage and analyze data. The course covers the following topics: descriptive statistics, sampling and probability, inference and hypothesis testing, OLS regression and associated topics such as model building, interaction effects and regression diagnostics. The course consists of weekly lectures and accompanying lab sessions led by a group of Graduate Teaching Assistants.

Assessment:
The course has two *formative* assessments—that is, work you submit for feedback but that does not affect your final score in the class—and one *summative* assessment—that is, work you submit for a marked grade that assesses your overall understanding of the material covered in this class. The two formative assessments are group assignments/problem sets, which will be submitted in Week 4 and Week 7 of Michaelmas, and on which you will receive feedback in lab during Week 5 and Week 8. The summative assessment is a take-home final exam, which covers all the material in the course, and occurs during Week 0 of Hilary Term. Unlike the assignments, the exam is based on your work alone, but you can reference class materials, additional readings, etc. as needed. During and after the exam, however, you should not correspond or share materials with your classmates, and any attempt to do so will be treated as academic dishonesty.

Class Materials, Set-up, and Installation:
The course relies heavily on either Quantitative Social Science: An Introduction (2018) by Kosuke Imai, or Quantitative Social Science: An Introduction in Tidyverse (2022) by Kosuke Imai and Nora Webb Williams. The latter is an updated version of the former, with updated instructions for the R (lab) content. Some copies of both are available in the social science library and other college libraries, but you may find it beneficial to purchase your own copy. If purchasing, we recommend Imai and Williams (2022), as it more closely follows our course content.

This class requires the use of R, a statistical programming language, and RStudio, a free software application that enables easier programming in R. If you have not already done so, you should make sure that you have the latest version of R installed, the latest version of RStudio, and you should make sure that you are able to access the internet during class (e.g., through eduroam) to download packages as needed. To download R, go to https://www.r-project.org/ and select the mirror (server) geographically closest to you. To download RStudio, go to https://posit.co/download/rstudio-desktop/, and select the installation for your system (Mac, Windows, or Linux). These programs should be installed before you come to the first day of lab.

During labs, you will be following along with code, so please have your laptop with you and be ready to work in pairs or small groups during in-class assignments. If you run into any errors during class (e.g., a package not loading or executing correctly), please try to resolve the error by looking it up on StackOverflow before requesting assistance in class.

We will use Canvas for some of the functions of class (e.g., distributing readings, lab materials, homeworks and answers, etc.), so please make sure you are correctly enrolled in the course and receiving notifications.

Finally, during lab, you will need access to the internet to download relevant materials), so we recommend making sure your Eduroam account is active and functioning correctly.

Other Policies:
If you are a parent and your childcare falls through, you are welcome to bring your child or infant to class provided they can be present without disrupting class. Similarly, if you are nursing, you are welcome to breastfeed in class. Please feel free to consume snacks or drinks during lecture, but reserve eating full meals for the breaks.

Please include "Intro Stats" in the subject line of your emails; if you do not, your email is likely to end up in the wrong folder and may be missed. We will try to respond to emails within 48 hours during the work week.

If you need disability-related accommodations in this class, and/or if you have emergency medical information that you wish to share, and/or if you need special arrangements to participate, please inform us as soon as possible via email or after the first class. For formal disability-related accommodations, you must also obtain a Student Support Plan, which will be sent by DAS to us.

Course Outline

Week 1 | Causality | Bernhard

Lecture: Scientific Method, Research Questions, Potential Outcomes, Experiments
Lab: R Basics
Required Readings:

Ware, William B., Ferron, John M., & Miller, Barbara M. (2013). Introductory Statistics: A Conceptual Approach Using R. Chapter 1. Available on the course Canvas website (overview—read to get a sense of what this class will cover).
Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Chapter 2 ('Causality')
OR
Imai, Kosuke., & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Chapter 2 ('Causality').

Recommended Readings:

Ismay, Chester, & Kim, Albert Y. (2023). Statistical inference via data science: A ModernDive into R and the tidyverse. Chapters 1, 3 and 4. Available at https://moderndive.com/ (covers similar topics on R code as the lab).

Week 2 | Measurement | Bernhard

Lecture: Conceptualization and Operationalization, Types of Variables, Summary Statistics, Measurement Error, Univariate Data Visualization
Lab: Merging Data and Data Visualization in R
Required Readings:

Paxton, Pamela. (2000). Women's suffrage in the measurement of democracy: Problems of operationalization. Studies in Comparative International Development, 35, 92-111. Available on SOLO.
Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Sections 3.1-3.4 ('Measurement' – 'Survey Sampling')
OR
Imai, Kosuke., & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Sections 3.1-3.4 ('Measurement' – 'Survey Sampling')

Recommended Readings:

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapter 3 ('Descriptive Statistics').
Ismay, Chester, & Kim, Albert Y. (2023). Statistical inference via data science: A ModernDive into R and the tidyverse. Chapter 2. Available at https://moderndive.com/ (covers similar topics on R code as the lab).

Week 3 | Relationships | Bernhard

Lecture: Relationships between Variables, Causal Diagrams, Correlations, Bivariate OLS
Lab: Implementing Summary Stats and Bivariate Regression in R
Required Readings:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Sections 3.6-3.7 ('Summarizing Bivariate Relationships,', 'Quantile-Quantile Plot'), Sections 4.1-4.2 ('Prediction' – 'Linear Regression')
OR
Imai, Kosuke., & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Sections 3.6-3.7 ('Summarizing Bivariate

Relationships,', 'Quantile-Quantile Plot'), Sections 4.1-4.2 ('Prediction' – 'Linear Regression').

Recommended Readings:

Ismay, Chester, & Kim, Albert Y. (2023). Statistical inference via data science: A ModernDive into R and the tidyverse. Chapter 5.1. Available at https://moderndive.com/ (covers similar topics on R code as the lab).
Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapters 7-9 ('Comparison of Two Groups' – "Linear Regression and Correlation')

Week 4 | Probability | Bernhard

Lecture: Randomization, Probability Theory, Central Limit Theorem, Sampling Strategies, Sampling Biases, Families of Distributions

Lab: Sampling, Standard Errors and Confidence Intervals in R

Required Readings:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Chapter 6 ('Probability')
OR
Imai, Kosuke., & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Chapter 6 ('Probability').

Recommended Readings:

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapter 4 ('Probability Distributions').

Assignment 1 (Covering Weeks 1-3 Material) Due by 11:59 pm on Friday

Week 5 | Analysis | Bernhard

Lecture: Hypothesis Testing, Confidence Intervals, Standard Errors, P-values, Simulations

Lab: Assignment 1 Reviewed; T-test and p-values.

Required Readings:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Sections 7.1-7.2 ('Estimation', 'Hypothesis Testing') OR Imai, Kosuke., & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Sections 7.1-7.2 ('Estimation', 'Hypothesis Testing')

Recommended Readings:

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapters 5-6 ('Statistical Inference: Estimation', 'Statistical Inference: Significance Tests').

Week 6 | Regression | Carella

Lecture: Understanding Linear Regression: Necessary Assumptions for OLS, Independence/BLUE, Multiple Regression

Lab: Multiple Regression in R and Regression Diagnostics

Required Readings:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Section 4.3.2 ('Regression with multiple predictors') and Section 7.3 ('Linear Regression Model with Uncertainty'). Review Sections 4.1 and 4.2 on bivariate OLS.

OR

Imai, Kosuke, & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Section 4.4.1 ('Regression with multiple predictors') and 7.3 ('Linear Regression Model with Uncertainty'). Review Sections 4.1–4.3 on bivariate OLS.

Recommended Readings:

Ismay, Chester, & Kim, Albert Y. (2023). Statistical inference via data science: A ModernDive into R and the tidyverse. Chapter 10. Available at https://moderndive.com/ (Lab)

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapter 11 up to section 11.3 inclusive.

Week 7 | Interactions | Carella

Lecture: Controlling, Interactions and Polynomial Terms, Multivariate Data Visualization

Lab: Understanding Interactions and Polynomial Terms in a Regression, Visualizing Model Estimates in R

Required Readings:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Section 4.3.3 ('Heterogeneous treatment effects')

OR

Imai, Kosuke, & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Section 4.4.2 ('Heterogeneous treatment effects')

Recommended Readings:

Cinelli, Carlos, Forney, Andrew, & Pearl, Judea. (2022). A crash course in good and bad controls. Sociological Methods & Research.

Hanretty, Chris. (2022). The effect of employment on attendance: A response to 'Identifying and understanding the drivers of student engagement'. Politics.

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Section 11.5 ('Interactions Between Predictors in Their Effects').

Assignment 2 (Covering Weeks 4-6 Material) Due by 11:59 pm on Friday

Week 8 | Nonlinear Regression | Carella

Lecture: Logistic Regression

Lab: Assignment 2 Reviewed, Implementing Logistic Regression in R

Required Readings:

James, Gareth, Witten, Daniela, Hastie, Trevor, & Tibshirani, Robert. (2021). An Introduction to Statistical Learning. Pages 129–139 (Chapter 4 up to 4.3 inclusive). Available online via SOLO.

Recommended Readings:

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapter 15 ('Logistic Regression: Modeling Categorical Responses').

Osborne, Jason W. (2014). Best practices in logistic regression. Chapters 1–-3. Available online via SOLO.

Final Exam Take-Home in Week 0 of Hilary Term