

# Introduction to Statistics for Political Science

Oxford University  
Department of Politics and International Relations  
Michaelmas 2024

**Instructors:** Rachel Bernhard | [rachel.bernhard@politics.ox.ac.uk](mailto:rachel.bernhard@politics.ox.ac.uk)  
Leonardo Carella | [leonardo.carella@nuffield.ox.ac.uk](mailto:leonardo.carella@nuffield.ox.ac.uk)

**GTA's:** Maksim Zubok | [maksim.zubok@nuffield.ox.ac.uk](mailto:maksim.zubok@nuffield.ox.ac.uk)  
Edmund Kelly | [edmund.kelly@hmc.ox.ac.uk](mailto:edmund.kelly@hmc.ox.ac.uk)  
Daniil Romanov | [daniil.romanov@sant.ox.ac.uk](mailto:daniil.romanov@sant.ox.ac.uk)  
Pablo Valdivieso Carrillo | [pablo.valdiviesocarrillo@politics.ox.ac.uk](mailto:pablo.valdiviesocarrillo@politics.ox.ac.uk)

**Lectures:** DPIR Lecture Theatre  
Wednesday 2.00-4.00pm

**Labs:** Lab 1a | XXXX | Seminar Room E | Friday 9am-10:30  
Lab 1b | XXXX | Seminar Room E | Friday 10:45am-12:15pm  
Lab 2a | YYYY | Lecture Theatre | Friday 9am-10:30  
Lab 2b | YYYY | Lecture Theatre | Friday 10:45am-12:15pm  
Lab 3a | WWWWW | Skills Lab | Friday 9am-10:30  
Lab 3b | WWWWW | Skills Lab | Friday 10:45am-12:15pm  
Lab 4a | ZZZZ | DPIR Boardroom 172 | Friday 9am-10:30  
Lab 4b | ZZZZ | DPIR Boardroom 172 | 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 and one summative assessments: (1) a formative take-home assignment ("problem set 1") due on Friday week 4 Michaelmas Term, covering material from lectures and labs from weeks 1-3; (2) a formative take-home assignment ("problem set 2") due on Friday week 8 of Michaelmas Term, covering material from lectures and labs from weeks 4-7; and (3) a summative final take-home exam due on Friday week 0 of Hilary Term, assessing the overall understanding of the material covered in the entire course. Formative assignments will be made available one week prior to the deadline; you can reference class materials, additional readings, etc. as needed. However, in any of the assessments you should not correspond or share materials with your classmates, and any attempt to do so will be treated as academic dishonesty. If you choose to use ChatGPT or similar tools in the problem sets, you are responsible both for noting clearly that you did so, and for any errors it generates (note that in previous years, reliance on ChatGPT caused students to miss at least one homework problem for which the answer was all but given in the class slides). You may not use ChatGPT or similar tools for the take-home final exam; as with asking classmates for materials or answers, this 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).

Jackman, S. (2008). Measurement, in: Box-Steffensmeier, J.M., Brady, H.E., Collier, D. (Eds.), *The Oxford Handbook of Political Methodology*. Oxford University Press. Available on SOLO.

### 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').

Blackwell, Matthew. (2013). A Framework for Dynamic Causal Inference in Political Science. *American Journal of Political Science*, 57(2), 504–520. Available on SOLO.

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')

Angrist, Joshua D., & Pischke, Jörn-Steffen. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. Chapters 2 and 3.

Huntington-Klein, Nick. (2021). *The effect: An introduction to research design and causality*. Chapman and Hall/CRC. Chapter 6 ('Causal Diagrams'). Available at <https://theeffectbook.net/>.

*Assignment 1 (Covering Weeks 1-3 Material) made available on Friday*

**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').

Freedman, David, & Pisani, Robert, & Purves, Roger. (2007.) *Statistics* (Fourth Edition). W.W. Norton & Company. Chapters 13-14 ('What Are the Chances?', 'More about Chance').

RECOMMENDED READINGS:

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

Freedman, David, & Pisani, Robert, & Purves, Roger. (2007.) *Statistics* (Fourth Edition). W.W. Norton & Company. Chapters 15-16.

*Assignment 1 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')

Freedman, David, & Pisani, Robert, & Purves, Roger. (2007.) Statistics (Fourth Edition). W.W. Norton & Company. Chapters 20-21 ('Chance Errors in Sampling', 'The Accuracy of Percentages').

RECOMMENDED READINGS:

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

Fisher, Ronald A. (1935). Design of Experiments. New York: Hafner. Chapter 1-2.

Rosenbaum, Paul R. (2020). Design of Observational Studies (Second Edition). Springer Nature.

## Week 6 | Linear Regression I | Carella

LECTURE: Bivariate ('Simple') Linear Regression Review, Multiple Regression, the R-squared, Multicollinearity, Omitted Variable Bias, Logic of 'Controlling'.

LAB: Multiple Regression in R, interpreting and exporting regression tables..

REQUIRED READINGS:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Sections 4.1-4.3.2 (up to 'Regression with multiple predictors' inclusive).

OR

Imai, Kosuke, & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Sections 4.1-4.4.1 (up to 'Regression with multiple predictors' inclusive).

Keller, Kayleigh. (2023). Introduction to Regression Analysis in R, available at: <https://www.kellerbiostat.com/introregression/> (Chapters 1, 2 and 8).

RECOMMENDED READINGS:

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

Gelman, A., Hill, J., & Vehtari, A. (2024). Regression and other stories, available at <https://users.aalto.fi/~ave/ROS.pdf>. (Chapters 6, 7 and 10 up to 10.2 inclusive).

Mine Cetinkaya-Rundel and Johanna Hardin. (2024). Introduction to Modern Statistics (2nd edition), available at <https://openintro-ims.netlify.app/>. Chapters 7 and 8.

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapter 9 (up to 9.4 inclusive), 10, and 11 (up to 11.3 inclusive).

Keller, Kayleigh (2023). Introduction to Regression Analysis in R, available at: <https://www.kellerbiostat.com/introregression/> (Chapters 3 and 8, more advanced material).

## Week 7 | Linear Regression II | Carella

LECTURE: Standard Error and Confidence Intervals of Regression Coefficients, Visualisation of Regression Results, Heteroskedasticity.

LAB: Visualising Model Predictions and Estimates in R, Regression Diagnostics.

REQUIRED READINGS:

Imai, Kosuke. (2018). Quantitative Social Science: An Introduction. Section 7.3 ('Linear Regression Model with Uncertainty').

OR

Imai, Kosuke, & Williams, Nora Webb. (2022). Quantitative Social Science: An Introduction in Tidyverse. Section 7.3 ('Linear Regression Model with Uncertainty').

Keller, Kayleigh. (2023). Introduction to Regression Analysis in R, available at: <https://www.kellerbiostat.com/introregression/> (Chapter 4, 'Inference on  $\beta$  in SLR').

RECOMMENDED READINGS:

Mine Cetinkaya-Rundel and Johanna Hardin. (2024). Introduction to Modern Statistics (2nd edition), available at <https://openintro-ims.netlify.app/>. Chapters 24 and 25.

Agresti, Alan, & Finlay, Barbara. (2009). Statistical methods for the social sciences. Chapter 9 (9.5-9.6), and 11 (section 11.5).

Keller, Kayleigh (2023). Introduction to Regression Analysis in R, available at: <https://www.kellerbiostat.com/introregression/> (Chapter 10, 'Inference in MLR', more advanced material).

*Assignment 2 (Covering Weeks 4-7 Material) made available on Friday*

Week 8 | Interactions and Transformations | Carella

LECTURE: Interactions, Polynomial Transformations, Log Transformations.

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').

Keller, Kayleigh. (2023). Introduction to Regression Analysis in R, available at: <https://www.kellerbiostat.com/introregression/> (Chapters 11 and 14).

RECOMMENDED READINGS:

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

Gelman, A., Hill, J., & Vehtari, A. (2024). Regression and other stories, available at <https://users.aalto.fi/~ave/ROS.pdf>. (Section 10.3).

Brambor, T., Clark, W., and Golder, M. (2006) "Understanding interaction models: Improving empirical analyses." *Political Analysis*, 14(1), 63-82.

Hainmueller, J., Mummolo, J., & Xu, Y. (2019). "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis*, 27(2), 163-192.

*Assignment 2 due by 11:59 pm on Friday*

*Final Exam Take-Home in Week 0 of Hilary Term*