Introduction to Quantitative Research Methods
Trinity College Dublin 2024/25

Tom Paskhalis

tom.paskhal.is

- **Module Code:** POP88162

- **Module Website:** tom.paskhal.is/POP88162

- **ECTS Weighting:** 10

- **Semester/Term Taught:** Semester 2 (Hillary Term)

- **Contact Hours:**
  One 2-hour lecture - Tuesday 09:00-11:00 in 2043 Arts Building
  One 1-hour workshop - Tuesday 13:00-14:00 in M20 Museum Building
  One 2-hour tutorial:

    – Group 1 - Thursday 09:00-11:00 in 1014 Arts Building

    – Group 2 - Thursday 16:00-18:00 in AP0.12 Aras an Phiarsaigh

  per week (11 weeks)

- **Module Coordinator:** Dr Tom Paskhalis (tom.paskhalis@tcd.ie)

- **Office Hours:** Thursday 11:00-13:00 in-person or online (booking required)

- **Teaching Assistants:**

    – Hannah Frank (frankh@tcd.ie)

    – Lucas Da Silva (ldasilva@tcd.ie)

Learning Aims

The goal of this module is to introduce students to the practice of data analysis at an elementary postgraduate level. More than ever before, political science research relies upon data — information about people, firms, nations, etc. that can be compiled, compared, and analyzed en masse. Political scientists analyze data with an eye to explaining the social world. Not all political scientists perform quantitative data analysis, but as empirical quantitative studies provide so much of our knowledge about politics and society, every student of the subject must now know at least a little about how it works.

Learning Outcomes

On successful completion of this module students should be able to:

- recognise different types of data;

- specify populations of observations and samples selected from such populations;

- develop statistical models of the social world;

- identify and summarise data useful for testing such models;

- carry out statistical hypothesis testing;

- interpret and critique quantitative results found in the political science literature;

Module Details

This module will consist of 3 parts: 2-hour lecture where we discuss approaches to empirical quantitative research and statistical methods, 1-hour workshop where we talk about using `R` for data analysis and 2-hour tutorial where you have a chance to have hands-on experience working with data using `R` and `RStudio`.

In the course of this module students will submit 3 assignments that are designed to test their knowledge of `R` and understanding of key statistical methods. Even more importantly than just being comfortable with working in `R`, the purpose of this course is to make you comfortable with design and execution of a quantitative study that answers a well-defined research question and is situated within some broader political science literature. To this end, you will be asked to write a research design and a final research paper.

Reading List

We will primarily be relying on the following core texts for this module:

- Alan Agresti. 2018. *Statistical Methods for the Social Sciences.* 5th ed. London: Pearson

- Ethan Bueno de Mesquita and Anthony Fowler. 2021. *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis.* Princeton, NJ: Princeton University Press

For more thorough treatment of causal inference and econometric models broadly refer to:

- Joshua D. Angrist and Jörn-Steffen Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect.* Princeton, NJ: Princeton University Press

- James H. Stock and Mark W. Watson. 2019. *Introduction to Econometrics.* 4th. London: Pearson

- Jeffrey M. Wooldridge. 2018. *Introductory Econometrics: A Modern Approach.* 7th ed. Boston, MA: Cengage

Some alternative textbooks on quantitative methods in social sciences using R that you might find useful:

- Andrew Gelman, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories.* Analytical Methods for Social Research. Cambridge: Cambridge University Press

- Kosuke Imai. 2017. *Quantitative Social Science: An Introduction.* Princeton, NJ: Princeton University Press

If you would like to consult additional texts on R and data analysis in R, you might find these books helpful:

- Norman Matloff. 2011. *The Art of R Programming: A Tour of Statistical Software Design.* San Francisco, CA: No Starch Press

- Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science.* 2nd. Sebastopol, CA: O'Reilly Media. https://r4ds.hadley.nz/

A classic book on research design (in its new edition), which introduces a lot of the general language and logic of inference that we will be using throughout this module:

- Gary King, Robert O. Keohane, and Sidney Verba. 2021. *Designing Social Inquiry: Scientific Inference in Qualitative Research, New Edition.* 2nd. Princeton, NJ: Princeton University Press

If you need a mathematics refresher or want to know more about the mathematical foundations of statistical procedures we discuss in this module I highly recommend this textbook:

- Jeff Gill. 2006. *Essential Mathematics for Political and Social Research.* Cambridge: Cambridge University Press

To learn more about data visualization and how to make your graphs look beautiful and convey information most accurately and efficiently see:

- Edward R. Tufte. 2001. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press. https://www.edwardtufte.com/tufte/books_vdqi

- Kieran Healy. 2019. *Data Visualization: A Practical Introduction.* Princeton, NJ: Princeton University Press

In addition, we will use a number of journal articles. Most journal articles will be freely available from the link included in the reading list (from campus computers). If this does not work (or if you are not on campus), search for the article via Trinity Stella Search (or Google Scholar) and log in to gain access.

Additional online resources:

- An Introduction to R

- Introduction to Econometrics with R

- Learning Statistics with R

- Learn R

- R Markdown Tutorial

- R Language Definition

- R Package Documentation

- YouTube:

    - First steps in R and RStudio
    - Intro to R
    - R Markdown

Data

Some useful links for obtaining data sets for quantitative analysis:

- Election and government information on 70+ countries - Political Parties, Presidents, Elections, and Governments Database

- Constituency-level election results for lower chamber (170+ countries) and upper chamber (24+ countries) legislative assemblies - Constituency-Level Elections Archive (CLEA)

- Democracy measures - Varieties of Democracy (V-Dem)

- Global beliefs and values - World Values Survey

- Global economic indicators - World Bank Data

- Parties, elections and cabinets in EU and OECD countries - ParlGov Project

- Party positions in Europe - Chapel Hill Expert Survey (CHES)

- Public attitudes, beliefs and behavioural patterns in Africa - Afrobarometer

- Public attitudes, beliefs and behavioural patterns in Europe - European Social Survey

- Public attitudes, beliefs and behavioural patterns in the EU - Eurobarometer

- Public attitudes, beliefs and behavioural patterns in Latin America - Latinobarómetro

- US elections - CQ Press Voting and Elections Collection

Software

In this class we will use `R` to work with data. `R` is free, open-source and interactive programming language for statical analysis. `RStudio` is a versatile editor for working with `R` code and data that provides a more intuitive interface to many features of the language.

Both R and RStudio are widely available for all major operating systems (Windows, Mac OS, Linux). You should install them on your personal computer prior to attending tutorials. Use these links to download the installation files:

- `R` - https://cran.r-project.org/

- `RStudio` - https://posit.co/download/rstudio-desktop/

Assessment Details

The final grade consists of the following parts (with corresponding weighting):

- Participation (10%)
  Tutorial attendance, RQ presentation

- 3 R assignments (5% each)

- Research design (15%)
  Approximately 1-2 pages and no more than 500 words (references excluded)

- Research paper (60%)
  Approximately 10 pages and no more than 5,000 words (references excluded)

The length of the research design assignment and the final research paper provided above should serve as a guide. There is a 10% leeway in the word count. That is any submission that falls 10% short or exceeds by 10% the word counts listed above will not be penalised. More detailed instructions for each assignment are provided in the accompanying Research Design Guidelines and Research Paper Guidelines documents.

In the research paper, each student will identify a research question and then answer it using quantitative analysis. Students must present their research question during one of the workshops after reading week or, otherwise, get an approval for their research question from the instructor.

All assignments should be submitted via Blackboard. Go to the "Assessment" section — you should be able to see all the assignments listed there. You will need to upload your assignments as DOC, DOCX, ODT or PDF documents. For R assignment submissions you should compile your R markdown into DOC or PDF document prior to submission.

Please make sure that you understand the submission procedure. Unexcused late submissions will be penalized in accordance with standard department policy. Five points per day will be subtracted until the Monday a week and a half after the deadline at which point the assignment is deemed to have failed.

All assignments are due by **08:59 Tuesday** prior to the start of the lecture. See module schedule summary below for the full list of due dates.

The final research paper will be due by **23:59 Tuesday, 22 April 2025**. See Research Paper Guidelines for further details.

Plagiarism

Plagiarism — defined by the College as the act of presenting the work of others as one's own work, without acknowledgement — is unacceptable under any circumstances. All submitted coursework must be **individual** and **original** (you should not re-use parts of a paper you wrote for another module, for example). You need to reference any literature you use in the correct manner. This is true for use of quotations as well as summarising someone else's ideas in your own words. When in doubt, consult with the lecturer before you hand in an assignment. Plagiarism is regarded as a major offence that will have serious implications. For more information on the College policy on plagiarism, please see avoiding plagiarism guide. All students must complete the online tutorial on avoiding plagiarism which can be found on this webpage.

Module Schedule

Week 1: Introduction

Required Readings:

- Ch 1 Agresti 2018

- George E. P. Box. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71 (356): 791–799. https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10480949

Additional Readings:

- Karl W. Broman and Kara H. Woo. 2018. "Data Organization in Spreadsheets." *The American Statistician* 72 (1): 2–10. https://doi.org/10.1080/00031305.2017.1375989

- John Gerring. 2017. "Qualitative Methods." *Annual Review of Political Science* 20 (1): 15–36. https://doi.org/10.1146/annurev-polisci-092415-024158

Week 2: Descriptive Statistics

Required Readings:

- Chs 2-3 Agresti 2018

Additional Readings:

- David Lazer et al. 2021. "Meaningful Measures of Human Society in the Twenty-first Century." *Nature* 595 (7866): 189–196. https://doi.org/10.1038/s41586-021-03660-7

- S. S. Stevens. 1946. "On the Theory of Scales of Measurement." *Science* 103 (2684): 677–680. https://www.science.org/doi/abs/10.1126/science.103.2684.677

Week 3: Probability Theory

Required Readings:

- Ch 4 Agresti 2018
- David Freedman. 1995. "Some issues in the foundation of statistics." *Foundations of Science* 1 (1): 19–39. https://doi.org/10.1007/BF00208723

Additional Readings:

- J. Neyman. 1977. "Frequentist Probability and Frequentist Statistics." *Synthese* 36 (1): 97–131. https://doi.org/10.1007/BF00485695
- Richard von Mises. 1941. "On the Foundations of Probability and Statistics." *The Annals of Mathematical Statistics* 12 (2): 191–205. https://doi.org/10.1214/aoms/1177731748

Week 4: Hypothesis Testing

Required Readings:

- Chs 5–6 Agresti 2018
- Ch 6 Mesquita and Fowler 2021

Additional Readings:

- Ronald A. Fisher. 1935. "The Principles of Experimentation, Illustrated by a Psycho-Physical Experiment." In *The Design Of Experiments.* Edinburgh: Oliver / Boyd. https://archive.org/details/in.ernet.dli.2015.502684
- Regina Nuzzo. 2014. "Scientific method: Statistical errors." *Nature* 506 (7487): 150–152. https://doi.org/10.1038/506150a

Week 5: Analysis of Proportions and Means

Required Readings:

- Chs 7–8 Agresti 2018

Additional Readings:

- Alan Agresti. 1992. "A Survey of Exact Inference for Contingency Tables." *Statistical Science* 7 (1): 131–153. https://doi.org/10.1214/ss/1177011454
- George E. P. Box and David R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (2): 211–243. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1964.tb00553.x

Week 6: Correlation

Required Readings:

- Chs 2–4 Mesquita and Fowler 2021

Additional Readings:

- Ch 9 Agresti 2018

- Joseph Lee Rodgers and W. Alan Nicewander. 1988. "Thirteen Ways to Look at the Correlation Coefficient." *The American Statistician* 42 (1): 59–66. https://doi.org/10.1080/00031305.1988.10475524

Week 8: Linear Regression I

Required Readings:

- Ch 9 Agresti 2018

- Ch 5 Mesquita and Fowler 2021

Additional Readings:

- Francis Galton. 1886. "Regression Towards Mediocrity in Hereditary Stature." *The Journal of the Anthropological Institute of Great Britain and Ireland* 15:246–263. https://doi.org/10.2307/2841583

Week 9: Linear Regression II

Required Readings:

- Ch 10 Mesquita and Fowler 2021

Additional Readings:

- Chs 10–11 Agresti 2018

- Gary King, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 347–361. https://gking.harvard.edu/files/making.pdf

Week 10: Linear Regression III

Required Readings:

- Ch 9, 11 Mesquita and Fowler 2021

Additional Readings:

- Ch 14 Agresti 2018

- Kosuke Imai and In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2): 467–490. https://imai.fas.harvard.edu/research/files/FEmatch.pdf

Week 11: Causation

Required Readings:

- Chs 13 Mesquita and Fowler 2021

- Paul W. Holland. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–960. https://www.tandfonline.com/doi/abs/10.1080/0162145 9.1986.10478354

Additional Readings:

- Luke Keele. 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23 (3): 313–335. http://doi.org/10.1093/pan/mpv007

- David Lewis. 1973. "Causation." *The Journal of Philosophy* 70 (17): 556–567. http://www. jstor.org/stable/2025310

Week 12: Logistic Regression

Required Readings:

- Ch 15 Agresti 2018

Additional Readings:

- Gary King and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9 (2): 137–163. https://gking.harvard.edu/files/0s.pdf

- David R. Cox. 1958. "The Regression Analysis of Binary Sequences." *Journal of the Royal Statistical Society: Series B (Methodological)* 20 (2): 215–232. https://rss.onlinelibrary.wiley. com/doi/pdf/10.1111/j.2517-6161.1958.tb00292.x

Module Schedule Summary

Week	Date	Lecture Topic	Workshop Topic	Tutorial Topic	Assignment Due
1	21 January	Introduction	R Overview	Getting Started with R	
2	28 January	Descriptive Statistics	Data Structures	Data & Variables	
3	4 February	Probability Theory	Probability Distributions	Distributions & Sampling	1 R Assignment
4	11 February	Hypothesis Testing	Data Frames	Data Frames & Plotting	
5	18 February	Analysis of Proportions & Means	Factor Variables	Cross Tabulation	
6	25 February	Correlation	Visualisations	Correlation	2 R Assignment
7	4 March	-	-	-	-
8	11 March	Linear Regression I	RQ Presentations I	Linear Regression I	
9	18 March	Linear Regression II	RQ Presentations II	Linear Regression II	
10	25 March	Linear Regression III	RQ Presentations III	Linear Regression III	3 R Assignment
11	1 April	Causation	RQ Presentations IV	Causation	
12	8 April	Logistic Regression	RQ Presentations V	Logistic Regression	Research Design