



**School of Politics and International Relations**

**Module Outline and Reading List**

**POL252  
POLITICAL RESEARCH**

**Semester B: Jan-March 2019**

**Lectures:  
Mondays 15-17 People's Palace PP1**

**Course convenor:  
Dr Javier Sajuria [j.sajuria@qmul.ac.uk](mailto:j.sajuria@qmul.ac.uk);  
Office: ArtsOne 2.29  
Drop-in hours Semester B: Mon 13-14; Wed 11-12**

**Seminar tutor:  
Dr Phillip Broniecki [p.broniecki@qmul.ac.uk](mailto:p.broniecki@qmul.ac.uk)  
Office: ArtsOne TBC  
Drop-in hours Semester B: Mon 14-16**

## **MODULE CONTENT**

This module provides essential research training in the use of quantitative methods in the political and social sciences, particularly for those interested in a future career in data analytics or in becoming critical consumers of quantitative research. The module caters for students from diverse academic disciplines and adopts a practical hands-on approach to learning, with tutor supported computer tutorials. The module covers descriptive statistics, data visualisation, data access, probability, sampling, hypothesis testing, inferential statistics and ends with an introduction linear regression. Students will be introduced to the R statistical software and work with data used in current academic research.

## **MODULE VALUE**

30 credits

## **MODULE OBJECTIVES**

By the end of the module, students will be able to:

- to familiarise you with a variety of quantitative research methods
- to provide you with essential research training skills in quantitative methods to support research within academia (appropriate to UG level) and beyond academia (leading to careers in a number of data-oriented sectors).
- to familiarise you with the R statistical language
- to enable you to design your own dissertation
- to enable you to interpret and critique published work using quantitative research methods

## **LEARNING OUTCOMES**

### *Knowledge*

By the end of this course you should be able to:

- demonstrate knowledge and understanding of key quantitative research techniques
- demonstrate an understanding of published work using quantitative research methods
- understand probability and statistical inference

### *Skills*

By the end of this course you should have developed:

- research skills, including the ability to select and use relevant resources effectively and to devise research questions appropriate for undergraduate research
- the ability to analyse and interpret published research using quantitative research methods

- technical competence in using R to perform a range of quantitative skills
  - conduct basic statistical tests
  - examine and visualise data using descriptive statistics
  - run and interpret linear regressions

## TEACHING ARRANGEMENTS

Each week there will be an introductory lecture followed by a computer tutorial. The lecture will last for two hours and the tutorial will last one hour. The lectures will introduce the ideas and issues relating to the various topics, while the tutorials will provide an opportunity to implement the techniques covered in the lectures. The first three tutorials provide an introduction to R, so students should expect a level of mismatch between the topics covered in the lectures and the ones covered in the tutorials.

Lectures will take place on Mondays 3-5pm in People's Palace PP1. Seminars will take place on Tuesdays, please check your timetable to make sure you attend the right seminar

## MODULE ASSESSMENT

Assessment is by coursework. Students will be expected to complete two projects, both of which should use original data analysis and the various methods and techniques covered in the quantitative part of the course. Projects should be focused on the analysis of data whilst locating their analysis within the context of addressing and answering a research question. The marker will look at the accurate and effective use of a range of methods taught in the course, the accuracy and presentation of graphs and tables, whether the tables are based on an original analysis of the data, whether the tables and graphs illustrate and advance ideas in the text and if the text accurately reports the results from the tables. Students are expected to conduct and provide evidence of the relevant statistical tests to verify their findings. Projects are pieces of academic writing and should make appropriate references to the literature and contain bibliographies. The usual penalties will apply to plagiarism.

The course has two marked components, a shorter midterm research project (worth 40% of the course mark) and a final take-home coursework (worth 60% of the course mark).

The midterm will review basic theory, test whether students have mastered basic data manipulation and model fitting in R, and also test whether students have done all the required reading and the assignments. The deadline for this assignment is **Sunday 17 February 2019**.

The take-home coursework will require students to address specific research or policy questions using real-world datasets; in other words, you will have to produce statistical analysis using R (<http://cran.r-project.org/>), and produce responses to the questions posed. The deadline for this assignment is **Sunday 7 April 2019**.

**Both pieces of coursework should be submitted electronically by midnight on the day of the deadline, via QM+.**

### *Optional Assignments*

At the end of each lecture section, an optional assignment will be available to students. These assignments DO NOT count toward the course mark! *They are optional* and intended to provide students formative feedback on the material presented. The optional assignments are available via the course's GitHub website

## **COURSEWORK SUBMISSION**

Students must:

- Upload your coursework by **midnight** on the day of the deadline to TurnItIn, in the relevant module area on QMPlus.
- Submit a single document electronically. The first page of your coursework should be the coversheet which is also available on QMPlus. You can download this Word document when you start typing your coursework; remember to save it as soon as you have completed the coversheet details, and begin your coursework on page 2.
- The school has a policy of anonymous marking and you should not type your name on your coursework or on the cover sheet; only use your QMUL student ID number.
- Please note work submitted without a coversheet will not be marked. It is your responsibility to submit a completed coversheet and submit on QMPlus by the deadline.
- Marks will be deducted if your coursework is not submitted on time.

**You must submit your coursework by the stated deadline. It is your responsibility to ensure that coursework is submitted correctly and fully. Failure to submit by the deadline will result in deductions for lateness.** Full details of submission policies can be found in the School's UG QMPlus website.

## **EXTENSIONS**

If you require an extension due to extenuating circumstances (EC), you must complete the relevant EC form and attach documentation to support your request. Completed forms and documentation should be handed into the Office. Full details can be found in the Student Handbook.

**Please make sure you check the Politics UG QMPlus website for the most updated version of the submission and extensions policy. In case of conflict with this handbook, the website information is the correct one.**

## **COURSE RESOURCES**

### *Readings*

Students are required to obtain the following textbooks for the module. These are online e-books and are free for students via their respective websites. Students can choose to obtain physical copies at their own cost.

- Diez et al. 2013. OpenIntro Statistics, 2nd Edition. <http://www.openintro.org/stat/textbook.php>.
- Lane et al. Online Edition. Introduction to Statistics. [http://onlinestatbook.com/Online\\_Statistics\\_Education.pdf](http://onlinestatbook.com/Online_Statistics_Education.pdf)

Students without any prior training in quantitative methods (statistics or econometrics) are recommended to complete background reading before each session. For background reading we recommend

- Kellstedt, Paul M., and Guy D. Whitten. 2013. *The Fundamentals of Political Science Research*. Cambridge: Cambridge University Press.

In addition to the main texts for the course, additional reading materials are required for some weeks. These are available online through QMPLus. It is expected that students will have read ALL the required reading *prior* to coming to lecture and seminar.

Items designated as ‘Further Reading’ are not required reading for the course. However, students often find these useful in the development of their dissertation proposals, or in providing some additional explanations on particularly difficult or relevant topics.

## R

Every quantitative social scientist (or data scientist more broadly) needs to know how to operate at least one piece of statistical software. In this course, we will be teaching you **how to use R through the RStudio IDE**. R is statistical software that allows one to manipulate data and estimate a wide variety of statistics. It is one of the fastest growing statistical software packages, one of the most popular data science software packages, and, importantly, it is open source (free!). In addition to the exercises that you will be asked to complete in seminars, you may also find the following tutorials helpful:

<http://data.princeton.edu/R/introducingR.pdf>

[http://cran.r-project.org/doc/contrib/Horton+Pruim+Kaplan\\_MOSAIC-StudentGuide.pdf](http://cran.r-project.org/doc/contrib/Horton+Pruim+Kaplan_MOSAIC-StudentGuide.pdf)

<http://cran.r-project.org/doc/contrib/usingR.pdf>

## QMPLUS and GitHub

All module materials, including a copy of this module outline, can be found on QMPlus. You should familiarise yourself with QMPlus as soon as possible as further information concerning this module and office hours will be posted there. To access QMPlus (on or off campus) go to <http://qmplus.qmul.ac.uk/>. You will require your QM computer access username and password.

You should also use QMPlus to upload the electronic version of your assignments.

If you are having problems accessing/using QMPlus support and information can be found on the following website –

<https://qmplus.qmul.ac.uk/login/index.php>.

Equally you can contact the School Office who may be able to offer assistance.

This module also uses a custom-made GitHub website that contains all the relevant materials for the tutorials. The address is.....

**In this website you will be able to check on the activities of each week and the assignments.**

**NOTE: If you have not completed your module registration properly your modules will not show on QMPlus. It is up to you to ensure you complete registration and check QMPlus regularly.**

### **BEFORE YOU TAKE THIS MODULE**

You should complete either one of the following online courses:

- Data Camp R tutorials <https://www.datacamp.com/courses/free-introduction-to-r>
- Codeschool: <http://tryr.codeschool.com/>

You can also use the following resource to learn more R during the module or have it as a reference text:

- *An Introduction to R*, available from <http://cran.r-project.org/doc/manuals/R-intro.pdf>
- If you find the above listed R tutorials and statistical material intuitive, we suggest you move to Joseph Adler's *R in a Nutshell* textbook available here:  
[http://web.udl.es/Biomath/Bioestadistica/R/Manuals/r\\_in\\_a\\_nutshell.pdf](http://web.udl.es/Biomath/Bioestadistica/R/Manuals/r_in_a_nutshell.pdf)

You should also download and install the latest version of RStudio (<http://www.rstudio.com>) and R (<https://cran.r-project.org>) on your computer.

### **PLAGIARISM**

QM defines plagiarism as 'presenting someone else's work as one's own irrespective of intention. Close paraphrasing, copying from the work of another person, including another student, using the ideas of another person, without proper acknowledgement or repeating work you have previously submitted without properly referencing yourself (known as 'self plagiarism') also constitute plagiarism.' – Regulations on Assessment Offences

Plagiarism is a serious offence and all students suspected of plagiarism will be subject to an investigation. If found guilty, penalties can include failure of the module, suspension or permanent withdrawal from Queen Mary.

It is your responsibility to ensure that you understand plagiarism and how to avoid it. The recommendations below can help you in avoiding plagiarism.

- Be sure to record your sources when taking notes, and to cite these if you use ideas or, especially, quotations from the original source. Be particularly careful if you are cutting and pasting information between two documents, and ensure that references are not lost in the process.
- Be sensible in referencing ideas – commonly held views that are

generally accepted do not always require acknowledgment to particular sources. However, it is best to be safe to avoid plagiarism.

- Be particularly careful with quotations and paraphrasing.
- Be aware that technology is now available at Queen Mary and elsewhere that can automatically detect plagiarism.
- Ensure that all works used are referenced appropriately in the text of your work and fully credited in your bibliography.
- Ensure that any direct quotations are placed within quotation marks and accompanied by a reference.

If in doubt, ask for further guidance from your adviser or module tutor. See your student handbook for further advice.

## MODULE OUTLINE

This course is taught rather differently to most other social science courses. Sessions are interactive and comprise a lecture followed by a hands-on computer lab session. The lectures introduce theories and concepts, while the lab sessions teach students how to apply these concepts in practice using R. Sessions are both challenging and cumulative, so it is extremely important that you make every effort to attend each week or you will fall behind and find it harder to follow. Please note that in order to use computer software during the tutorials, they will be held in a computer lab.

Readings are provided for each individual week. The course does not assume any prior knowledge of statistics or R, and the course begins at an introductory level, although it advances quickly so students should allow the necessary preparation time each week to enable them to keep pace. Students are expected to do sufficient reading to understand the concepts introduced each week, and more importantly, to practise data analysis between sessions to make sure that they are confident about applying what they have learned. R can be used on university computers and above you can find information on where to obtain a copy for your own computers (for free!).

### Week 1

#### **Lecture: Understanding Data**

After providing an overview of the aims, learning objectives and practical arrangements for the module, the lecture moves on to take a brief look at the notion of variables and how they are measured, frequency distributions and ways of describing the central tendency and the dispersion of a variable.

#### **Tutorial: Introduction to R**

The first week provides an introduction to the statistical software R, which will be used in all subsequent tutorials.

Students will learn how to download and install R, enter simple data into R, how to load data sets from other statistical packages into R and how to use the R environment for simple calculations.

#### *Required Reading*

- Diez et al. Sections 1.1-1.2
- Lane et al. Chapters 1 and 3

### *Further Reading*

- Garner, Roberta. 2010. *The Joy of Stats: A Short Guide to Introductory Statistics in the Social Sciences*. Ontario: University of Toronto Press. (especially the “Math Refresher” section)
- Rowntree, D. 1991. *Statistics without Tears: A Primer for Non-Mathematicians*. London: Penguin.
- Salkind, Neil J. 2004. *Statistics for People Who Think They Hate Statistics*. London: Sage.
- Tufte, Edward. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press LLC.
- Yau, Nathan. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Indianapolis, IN: Wiley Publishing, Inc.

## **Week 2**

### **Lecture: Examining data 1**

This session explores a dataset to identify and describe different types of variables using R. Various techniques for looking at variable distributions, including table and basic univariate graphs.

### **Tutorial: Describing Data 1**

The practical provides an introduction to univariate (one variable) descriptive statistics and graphs in R, including frequency tables, histograms and boxplots.

### *Required Reading*

- Diez et al. Sections 1.6 – 1.7
- Lane et al. Chapter 2

### *Further Reading*

- Babbie, E. 1997. *The Practice of Social Research*. Belmont, CA: Wadsworth Publishing Company.
- Weisberg, Herbert. 1991. *Central Tendency and Variability*. Series: Quantitative Applications in the Social Sciences. Thousand Oaks: Sage.
- Pennings, Paul, Hans Keman, and Jan Kleinnijenhuis. 1999. *Doing Research in Political Science: An Introduction to Comparative Methods and Statistics*. London: Sage.
- McClave, James T., and Terry Sincich. 2003. *A First Course in Statistics*, 8<sup>th</sup> ed. New Jersey: Prentice Hall.

## **Week 3**

### **Lecture: Examining Data 2**

This session explores a dataset to identify and describe different types of variables using R. Various techniques for looking at variable distributions, including tables, graphs and summary statistics are considered. We also address access to data and data management such as how to recode a variable.

### **Tutorial: Describing data 2**

This practical explores descriptive bivariate statistics and graphs using R, including two-way tables, bar charts, line graphs and scatterplots.

### *Required Reading*

- Lane et al. Chapter 4 (up to p. 170)



## **Week 4**

### **Lecture: Probability and the normal distribution**

The basics of probability theory are introduced as well as concepts of probability distributions, the normal curve, and sampling distributions of means.

### **Tutorial: Sampling distribution of the mean**

This session introduces sampling distributions using a simple simulated data set in R.

#### *Required Reading*

- Diez et al. Chapter 3 (up to p. 128) & Chapter 4 (up to p. 164)
- Lane et al. Chapters 7 and 9

## **Week 5**

### **Lecture: Confidence intervals and significance**

The session explores the ability to generalise the findings from analysis of sample data to the wider population (inference). The theory of hypothesis testing, confidence intervals and statistical significance are introduced.

### **Tutorial: Testing for sampling error**

This practical follows on from the previous week using the simulated data to derive confidence intervals.

#### *Required Reading*

- Diez et al. Sections 4.2-4.3
- Lane et al. Chapters 10 - 11

#### *Further Reading*

- DeGroot, Morris, and Mark Schervish. 2002. *Probability and Statistics*. Massachusetts: Addison-Wesley.

## **Week 6**

### **Lecture: Measures of difference**

This lecture considers statistical hypothesis tests for the difference in two means from paired and independent samples and the difference in two categorical variables.

### **Tutorial: Tests of significance**

The session explores the use of appropriate use of test of significance for a mean.

#### *Required Reading*

- Diez et al. Chapter 5 and sections 6.3 and 6.4
- Lane et al. Chapters 12 and 17

## **Week 7**

**Reading week – no class**

## **Week 8**

### **Lecture: Exploring relationships between interval variables**

This lecture starts with a look at graphical approaches of exploring the relationship between interval variables using scatterplots introduced in lecture 2 before moving on to look at measures of correlation and associated statistical tests and how to interpret them.

### **Tutorial: Exploring relationship between categorical variables**

We cover how to test the statistical significance of two-way samples using a simple Chi Square test.

#### *Required Reading*

- Diez et al. Section 7.1
- Lane et al. Chapter 4

#### *Further Reading*

- Field, A., Miles, J. and Field, Z. (2012) *Discovering Statistics Using R*. London: Sage. Chapter 6.

## **Week 9**

### **Lecture: The simple linear regression model**

This lecture introduces the theory and practice of simple linear regression and how to interpret the output. The components of a simple linear regression model are described and explained before looking at how to interpret the output and checking the assumptions of the method are met.

### **Practical: Correlation and linear regression**

The session first covers how to run and interpret of two statistical tests of correlation (Pearson's  $r$  and Spearman's  $\rho$ ) and then how to design, run, and interpretation of a simple linear model.

#### *Required Reading*

- Diez et al. Sections 7.2-7.4
- Lane et al. Chapter 14 (up to p.483)

#### *Further Reading*

- Field et al. Chapter 7 (up to p. 260)
- Berry, William D., and Stanley Feldman. 1985. *Multiple Regression in Practice*. London: Sage.
- Wooldridge, Jeffrey. 2012. *Introductory Econometrics: A Modern Approach*. New York: South Western Cengage Learning. (Chapter 6)
- Achen, Christopher. 1982. *Interpreting and Using Regression*. Series: Quantitative Applications in the Social Sciences, No. 29. Thousand Oaks, CA: Sage.

## **Week 10**

### **Lecture: Assumptions of linear regression**

The purpose of this lecture is to explore the basic assumptions underlying the multiple linear regression model such as collinearity, outliers/leverage and correlated residuals.

### **Tutorial: Testing the assumptions of multiple linear regression**

This tutorial enables students to assess their linear regression models by testing for the basic assumptions underlying this statistical method.

#### *Required Reading*

- Diez et al. Section 8.3
- Lane Chapter 7 (up to p. 483)

#### *Further Reading*

- Field et al. Section 7.7 and 7.9

## **Week 11**

### **Lecture: Bonus topic, social media and web research**

We will explore basic methods for doing social media data gathering and analysis, using most of the techniques learnt during the year.

### **Seminar: Techniques for social media data.**

We will explore basic techniques for text analysis and social network analysis using social media data.

#### *Required Reading*

- Andreas Jungherr & Yannis Theocharis (2017) The empiricist's challenge: Asking meaningful questions in political science in the age of big data, *Journal of Information Technology & Politics*, 14:2, 97-109, DOI: 10.1080/19331681.2017.1312187

#### *Further Reading*

- Sajuria, J., vanHeerde-Hudson, J., Hudson, D., Dasandi, N. and Theocharis, Y., 2015. Tweeting alone? An analysis of bridging and bonding social capital in online networks. *American Politics Research*, 43(4), pp.708-738.

## **Week 12**

### **Lecture: Putting it all together**

Data analysis and evidence-based policy making and research

### **Tutorial: Exercises in preparation for final assessment.**