UCL DEPARTMENT OF POLITICAL SCIENCE
SCHOOL OF PUBLIC POLICY

Module Name: Introduction to Quantitative Research Methods

Module Code: PUBLG100A/B

Teaching: 20 hours of lectures, 10 hours of seminars

Credits: 15

Assessment: Midterm take-home coursework (25%) and take-home coursework (75%)

Essay Deadline/s: Midterm: 8th November.

Lecturer: Dr Jack Blumenau and Philipp Bronieki

Office Hours: JB: 13.00 – 15.00, Monday
PB: 15.00 – 17.00, Monday

Please note that it is not possible for students to register on both Introduction and Advanced Quantitative Methods. Students must select PUBLG100A/B Introduction to Quantitative Methods or PUBLG088 Advanced Quantitative Methods

USEFUL LINKS

PG Student Intranet
http://www.ucl.ac.uk/political-science/intranet/pg

Lecture and Seminar Times:
www.ucl.ac.uk/timetable

Extenuating Circumstances
http://www.ucl.ac.uk/political-science/intranet/pg/pastoral/extenuating_circumstances

Essay Extensions
http://www.ucl.ac.uk/political-science/intranet/pg/essays/extensions

Penalties for Late Submission and Overlength Essays
http://www.ucl.ac.uk/political-science/intranet/pg/essays/lateness_word_penalties

Essay Submission Information
http://www.ucl.ac.uk/political-science/intranet/pg/essays/submission_return

Examinations
http://www.ucl.ac.uk/political-science/intranet/pg/exams-dissertations

Plagiarism and TurnItIn
http://www.ucl.ac.uk/political-science/intranet/pg/policies/plagiarism

PUBLG100 – Introduction to Quantitative Research Methods
Autumn 2017

Course Tutors

Dr Jack Blumenau
Email: j.blumenau@ucl.ac.uk
Office: B.13, 29/30 Tavistock Sq.
Office Hours: 13.00 – 15.00, Monday

Philipp Broniecki
Email: philipp.broniecki.14@ucl.ac.uk
Office: 1.01, 31 Tavistock Sq.
Office Hours: 15.00 – 17.00, Monday

Introduction

This course is designed to introduce you to and help you become familiar with data science which is critical to your development as a social scientist. The introductory course has two primary aims. First, students will be introduced to statistical models that researchers and policymakers use in answering social, political and economic questions. Second, the course will equip students to use one or more of the discussed techniques in their MSc dissertation. By the end of the course, students should be able to *understand* the quantitative tools employed in political, social, and economic research; to *perform* data analysis using statistical software *R* and *interpret* results; and to fruitfully *employ* data science techniques in their dissertation research and in subsequent careers.

This module (or the Advanced Quantitative Methods module) is required of all students pursuing an MSc from the School of Public Policy, including degrees in Democracy and Democratization, European Public Policy, Global Governance and Ethics, International Public Policy, Public Policy, and Security Studies.

Course Format

This course will be taught using a combination of lectures and lab sessions. Lectures are two-hours in length. The lectures are designed to introduce students to the topics outlined in the course syllabus and to detail topics covered in the assigned readings. Lab sessions are designed to provide students the opportunity to get 'hands on' experience with the material and the statistical software— R (http://cran.r-project.org/)—and are capped at 15 students (depending on class size, etc.). Because the material in the course builds on previous weeks, students should plan on attending seminars from week 1 and attend consistently. Attendance of both lectures and lab sessions are required of all students.

Course Meeting Information

The course is delivered in both lecture and seminar format. For the most up to date information, see the online timetable at: www.ucl.ac.uk/timetable

Lectures

There are two lectures for this module, one section taught by Jack (PUBLG100B) and the other by Philipp (PUBLG100A). The material presented in each of the sections is exactly the same. However, your work will be assessed by the tutor to which you have been assigned. Students MUST attend the lecture slots they have been assigned.

PUBLG100A (Jack Blumenau): Venue TBD
PUBLG100B (Altaf Ali): Venue TBD

Seminars
Seminars (lab sessions) are primarily devoted to understanding how to apply the data science techniques discussed in the lecture to specific research and policy problems using *R*. Note that you will be required to use *R* to complete your final exam.

You will automatically be allocated to a seminar group. Attendance of seminars (lab sessions) is mandatory.

Seminar	Day	Time	Location	Tutor
1	Monday	11:00-12:00	25 Gordon Street, 105*	Shirley Dorchin
2	Monday	11:00-12:00	Chadwick Building, 2.23	Anastasia Ushakova
3	Monday	11:00-12:00	26 Bedford Way, 316	Sofia Collignon
4	Monday	11:00-12:00	26 Bedford Way, G.11	Philipp Shroeder
5	Monday	11:00-12:00	Cruciform B1.15A	Baris Ari
6	Monday	12:00-13:00	Chadwick Building, 2.23	Philipp Broniecki
7	Monday	12:00-13:00	26 Bedford Way, G.11	Philipp Shroeder
8	Monday	12:00-13:00	26 Bedford Way, 316	Sofia Collignon
9	Monday	12:00-13:00	25 Gordon Street, 105*	Jack Blumenau
10	Monday	12:00-13:00	Birkbeck, Mallet Street, 458	Baris Ari
11	Monday	13:00-14:00	Chadwick Building, 2.23	Anastasia Ushakova
12	Monday	13:00-14:00	26 Bedford Way, G.11	Shirley Dorchin
13	Monday	13:00-14:00	Foster Court, B29	Philipp Shroeder
14	Monday	13:00-14:00	26 Bedford Way, 316	Sofia Collignon
15	Monday	13:00-14:00	DMS Watson, G15	Baris Ari
16	Monday	14:00-15:00	26 Bedford Way, 316	Shirley Dorchin
17	Monday	14:00-15:00	26 Bedford Way, G.11	Anastasia Ushakova
18	Monday	17:00-18:00	26 Bedford Way, 316	Sofia Collignon
19	Monday	17:00-18:00	DMS Watson, G15	Shirley Dorchin
20	Monday	17:00-18:00	Archaeology 117	Philipp Shroeder
21	Monday	17:00-18:00	Cruciform, B1.15A	Baris Ari
22	Monday	17:00-18:00	20 Bedford Way, 427	Anastasia Ushakova

* Different rooms some weeks. See https://timetable.ucl.ac.uk for full details

Course Assessment
Students must pass this course to successfully complete the MSc degree. The course has two marked components, a short midterm take-home coursework (worth 25% of the course mark) and a final take-home coursework (worth 75% 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 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.

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 webpage.

Course Resources

Readings
Students are required to obtain the following textbook for the module.

- Stock, James H., and Mark W. Watson. 2014. *Introduction to Econometrics.* London: Pearson.

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.

You may also wish to purchase this textbook if you feel you will frequently refer to background reading. Kellstedt & Whitten provides an easy introduction to the same topics as in Stock & Watson. Some may find it useful to start with Kellstedt & Whitten and then read the corresponding chapters in Stock & Watson. It is important to note that Kellstedt & Whitten by itself is not sufficient to pass this module.

Both books are available in the library. However, given the large class size, we strongly recommend that students purchase personal copies of these texts.

In addition to the main texts for the course, additional reading materials are required for some weeks. These are available online through the course's electronic reading list (http://readinglists.ucl.ac.uk/modules/publg100.html). 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.

Moodle
We will make extensive use of UCL's virtual learning platform, Moodle. Students will be automatically enrolled in Moodle for the course to which they have been assigned, either:
100A (https://moodle.ucl.ac.uk/course/view.php?id=14165) or
100B (https://moodle.ucl.ac.uk/course/view.php?id=19248)

Piazza

We are using a service called *Piazza.com* to manage communications for this course. Piazza can be accessed via a link in the "Discussion Forums" section of the course's Moodle page. This is a much more efficient mode of communication than e-mail because it allows you to answer each other's questions, which will be much faster than waiting for a response from us, and for the entire class to see our responses, ensuring that we do not answer the same question multiple times over e-mail. Note that we primarily expect you to use Piazza for student-to-student communication, meaning that you should be attempting to answer each other's questions. We will check in with the discussions on Piazza each week to steer conversations in the right direction, but will not regularly check and answer each question.

In addition to Piazza, we will of course be happy to answer questions either before or after lecture or during our office hours.

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*. *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/usingR.pdf

Before you take this module

You should complete the following:

- Data Camp R tutorials https://www.datacamp.com/courses/free-introduction-to-r
- Codeschool: http://tryr.codeschool.com/
- *An Introduction to R*, available from http://cran.r-project.org/doc/manuals/R-intro.pdf
- Kellstedt, Paul M. and Guy D. Whitten (2013). *The Fundamentals of Political Science Research*, 2nd edition, Cambridge University Press, Chapters 5-11. If you haven't had any quantitative methods (statistics or econometrics) training in your undergraduate studies, you should go through these chapters as necessary background material for the module. You can also treat it as a refresher of basic statistics if you've done it previously.
- If you find the above listed R tutorials and statistical material in Kellstedt & Whitten (chapters 5-11) 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

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.

Course Outline

Week 1 – Introduction to Quantitative Analysis
Introduction to Quantitative Methodology; Causality; Research Design, Levels of measurement, central tendency and measures of dispersion.

Required Reading
- Stock and Watson (ch 1)
- Lohr, Steve. 2009. "For Today's Graduate, Just One Word: Statistics." *New York Times*. Available at: http://www.nytimes.com/2009/08/06/technology/06stats.html
- Lohr, Steve. 2012. "The Age of Big Data." *New York Times*. Available at: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Background Reading
- Kellstedt and Whitten (chs 1-4)

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 – Sampling and Distributions
Probability and random variables, useful distributions, expected values, random sampling, and sampling distributions.

Required Reading
- Stock and Watson (ch 2)

Background Reading
- Kellstedt and Whitten (ch 5 and 6)

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*, 8th ed. New Jersey: Prentice Hall.

Week 3 – T-test for Difference in Means and Hypothesis Testing
Significance tests; difference in means

Required Reading
- Stock and Watson (ch 3.2-3.3)

Background Reading
- Kellstedt and Whitten (ch 7)

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

Week 4 – Bivariate linear regression models
Linear regression with one explanatory variable; comparing linear regression and t-test for means; uncertainty, goodness of fit, prediction

Required Reading
- Stock and Watson (ch 4 & 5.1-5.3)

Background Reading
- Kellstedt and Whitten (ch 8.1-8.4)

Further Reading
- Levin, Jack, James Fox, and David Forde. 2009. Elementary Statistics in Social Research, 11th edition. (International ed. Pearson/Allyn and Bacon. (chs 9 and 12)

Week 5 – Multiple linear regression models (I)
Linear regression with multiple explanatory variables; control variables, F-test for joint hypothesis testing

Required Reading
- Stock and Watson (6.2-6.6 & 7.1-7.3 & 7.5)

Background Reading
- Kellstedt and Whitten (ch 9)

Further Reading
- Bray, James H., and Scott Maxwell. 1985. *Multivariate Analysis of Variance*. Series: Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

Week 6 – Multiple linear regression models (II)

Dummy variables as explanatory factor; interactions with dummy variables; interaction of two continuous variables, fixed-effects.

Required Reading

- Stock and Watson (ch 8.3)

Background Reading

- Kellstedt and Whitten (ch 10.1-10.3)

Further Reading

- 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 7 – Regression Assumptions and Violations of Assumptions

Heteroscedasticity, omitted variable bias, serial correlation

Required Reading

- Stock and Watson (5.4-5.6 & 6.1 & 9)

Background Reading

- Kellstedt and Whitten (ch 8.5)

Further Reading

- Berry, William D. 1993. *Understanding Regression Assumptions*. Series: Quantitative Analysis in the Social Sciences. Thousand Oaks: Sage.
- Fox, John. 1991. *Regression Diagnostics: An Introduction*. Series: Quantitative Applications in the Social Sciences. Thousand Oaks: Sage Publishers.

Week 8 – Binary models: Logit (I)

Discrete and Limited Dependent Variables; Probability and odds; Logistic regression; Coefficient interpretation; predicted probabilities

Required Reading

- Stock and Watson (ch 11)
- Long, Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications. [Chapter 3-3.5]

Background Reading

- Kellstedt and Whitten (ch 11.1-11.2)

Further Reading
- Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons.
- Hosmer, D., and S. Lemeshow. 1989. Applied Logistic Regression. New York: John Wiley. (Chs. 1-3)
- Aldrich, John H., and Forrest D. Nelson. 1984. *Linear Probability, Logit and Probit Models*. Series: Quantitative Applications in the Social Sciences, 45. London: Sage.

Week 8 – Binary models: Logit (II)
Maximum likelihood estimation; Inference for the Logit model; Z-scores; Likelihood-ratio tests; Goodness of fit.

Required Reading
- Stock and Watson (ch 11)
- Long, Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications. [Chapter 3-3.5]

Background Reading
- Kellstedt and Whitten (ch 11.1-11.2)

Further Reading
- Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons.
- Hosmer, D., and S. Lemeshow. 1989. Applied Logistic Regression. New York: John Wiley. (Chs. 1-3)
- Aldrich, John H., and Forrest D. Nelson. 1984. *Linear Probability, Logit and Probit Models*. Series: Quantitative Applications in the Social Sciences, 45. London: Sage.

Week 10 – Putting it all together
Data analysis and evidence-based policy making and research

Required Reading
- Kellstedt and Whitten (chs 1-4, 12)

Further Reading
- King, G., R. Keohane, S. Verba (1994). *Designing Social Inquiry*. Princeton University Press.