GRAD-C5: Statistics 1
Philipp Broniecki

1. General information

Class time	Monday 10 – 12h (Group A), 14 – 16 h (Group B)
Venue	Forum
Instructor	Philipp Broniecki
Instructor's office	
Instructor's e-mail	philippbroniecki@gmail.com
Assistant	Name: Andrea Derichs-Carlin Email: adjunctsupport@hertie-school.org Phone: +49 30 259 219 312 Room: 2.55
Instructor's Office Hours	See information to "Piazza" (p.2)

Instructor Information:
Philipp Broniecki researches legislative politics with a focus on the European Union and quantitative methods. His work relates to the effects of institutional design on decision-making. He holds a B.A. from the Free University of Berlin, an M.A. from the University of Mannheim, and is a PhD candidate at the Department of Political Science, University College London.

2. Course Contents and Learning Objectives

This course is an introduction to data science. The course has three primary aims. First, students will be introduced to the logic of quantitative research design. Second, we will cover statistical models that scientists and policy-makers use to answer social science questions. Third, students will acquire the necessary skills to conduct their own quantitative research projects. No prior statistical knowledge is assumed. We will use the statistical software R.

The course is delivered in both lecture and seminar format. Attendance in both lecture and seminar (lab session) is mandatory. Students are encouraged to stay on top of the readings and to make sure they complete weekly homework assignments and review the material covered in the lab sessions. The course is highly cumulative. Students should, therefore, ask questions online, in office hours, in class, and work together with their peers on assignments.

Material for this course will be available on our course website which we announce in the first lecture. The syllabus and weekly assigned readings may be updated over the course of the semester. The course website will contain up-to-date information.

Course delivery and preparation:

Lectures are 2 hours long and labs 1.5 hours. Lectures 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 with the opportunity to get 'hands-on' experience with the material and the software. Our labs are capped at 15 students. Students will be assigned to a lab session at random.

While we do not assume prior knowledge of statistics or computer coding, we encourage students to prepare for this class by taking the following steps.

Step 1: Download R. R is the statistical software we will use in this class. The latest version is available here: https://www.r-project.org/

Step 2: Download RStudio (https://www.rstudio.com/). RStudio is an editor (IDE) for the language R and it is the software, we will use all the time. We will write all our code in RStudio.

Step 3: Complete the free online course 'Introduction to R' on datacamp.com.

Piazza:

We are using a service called Piazza.com to manage communications for this course. Piazza is an online forum and can be accessed via a link on moodle (we will go over this in the first lecture). Piazza is a much more efficient mode of communication than e-mail. It allows you to see each other's questions and to answer each other's questions. This will be faster than waiting for a response from us, allow shy students to profit from questions asked by their peers, and ensure that we do not answer the same questions multiple times. 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 Piazza to steer conversations in the right direction.

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

3. Grading and Assignments

The course has three marked components. First, a midterm exam (worth 30% of the course mark). Second, a final take-home exam (worth 40% of the course mark). Third, weekly homework assignments (worth 30% of the course mark).

The midterm exam will be taken in class and contain questions on the material covered up until the midterm. You are allowed to take a calculator but you cannot take a computer or a laptop. The exam will take 1.5 hrs.

In the take-home exam you are given data (via Moodle) and asked to apply the skills you have learned. The exam will contain three parts. (1) Review questions, (2) data section 1, (3) data section 2. In the data sections you are presented with a research question, given data and asked to come up with a solution. The word limit for the exam is 4000 words.

Homework assignments will be posted on Moodle in sessions 2-11 and they are due on Fridays 11.59 pm (submission via Moodle). Late submissions cannot be accepted because we publish the grades and solutions at midnight. The assignments will be multiple choice. The lowest grade on a homework assignment—including a zero for one that is not submitted—will be dropped with no questions asked. Use this privilege wisely; it is intended for unforeseen circumstances, accidents, and illnesses. Each weekly assignment is worth 3.33% of the final grade.

Grade appeals should be submitted to the appropriate TA specifying the question(s) in doubt and grounds for the appeal.

Composition of Final Grade:

Midterm exam	Deadline: 22.10.2018 10-12 h (group A); 14-16 h (group B)	Handed out in class and submission in class	30%

		(calculators allowed, no computers/books/scripts)	
Final take-home coursework	Deadline: 17.December 2018, 23:59h	Submit by: exam question will be posted on Moodle on 10.12.2018 and has to be upload on 17.102018 until 23:59 on Moodle	40%
Homework assignments	Deadline: weekly from session 2-11 at 11.59 p.m.	Submit by: Moodle	30%

Late submission of assignments:
For each day the assignment is turned in late, the grade will be reduced by 10% (e.g. submission two days after the deadline would result in 20% grade deduction).

Attendance: Students are expected to be present and prepared for every class session. Active participation during lectures and seminar discussions is essential. If unavoidable circumstances arise which prevent attendance or preparation, the instructor should be advised by email with as much advance notice as possible. Please note that students cannot miss more than two sessions. For further information please consult the Examination Rules §9.

Academic Integrity: The Hertie School of Governance is committed to the standards of good academic and ethical conduct. Any violation of these standards shall be subject to disciplinary action. Plagiarism, deceitful actions as well as free-riding in group work are not tolerated. See Examination Rules §15.

4. General Readings

Students are strongly recommended to obtain the following textbook for the module:

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

While not required for this course, we also recommend the following books for a broader understanding of data science:

- Kellstedt, Paul M., and Guy D. Whitten. 2013. *The Fundamentals of Political Science Research. Cambridge: Cambridge University Press.*
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2015. *Mastering `Metrics. The Path from Cause to Effect.* Princeton: Princeton University Press.
- Imai, Kosuke. 2017. *Quantitative Social Science. An Introduction.* Princeton: Princeton University Press.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R.* New York: Springer. Available online at https://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research.* Princeton: Princeton University Press.

Students who want to gain a deeper understanding of probability theory may read:

- DeGroot, Morris H., and Mark. J. Schervish. 2012. *Probability and Statistics.* Fourth Edition. Pearson Education.

In addition to the main textbooks, additional reading materials are required for some weeks. There are available online through the course's electronic reading list. It is expected that students read the required texts prior to the lectures. Re-reading after the lecture is recommended as well.

5. Session Overview

Session	Session Date	Session Title
1	03.09.2018	Introduction: Measurement, Central Tendency, Dispersion, Validity, Reliability
2	10.09.2018	Research Design: Counterfactuals, Forming Hypotheses, Internal & External Validity
3	17.09.2018	Sampling and Distributions
4	24.09.2018	T test for the Difference in Means and Hypothesis Testing
5	01.10.2018	T tests for the Difference in Means, Confidence Intervals and Revision – please bring substantial questions
6	08.10.2018	Bivariate Linear Regression Models
7	15.10.2018	Multiple Linear Regression Models (I)
Mid-term Exam Week: 22-26 October 2018 – tba		
8	29.10.2018	Multiple Linear Regression Models (II)
9	05.11.2018	Regression Assumptions and Violations of Assumptions
10	12.11.2018	Panel Data Structures: Fixed Effects Models
11	19.11.2018	Binary Outcomes: The Linear Probability Model
12	26.11.2018	Putting It All Together or RCTs: Randomized Controlled Trials (will be announced)
Final Exam Week: 10-14 December 2018 – no class		

6. Course Sessions and Readings

Session 1: 03.09.2018 **Introduction: Measurement, Central Tendency, Dispersion, Validity, Reliability**	
Learning Objective	Motivation for quantitative methods, what are random of variables, how do we measure concepts, how do we describe our data
Required Readings	Stock & Watson Ch 1
Optional Readings	Imai Ch 3; Kellstedt & Whitten Ch 5

Session 2: 10.09.2018 **Research Design: Counterfactuals, Forming Hypotheses, Internal & External Validity**

Learning Objective	Understanding of questions that can be asked with quantitative methods; how to approach answering a question of interest and how to evaluate the approach
Required Readings	Kellstedt & Whitten Chs 1--4 (will be uploaded on moodle)
Optional Readings	Gary Taubes, "Do We Really Know What Makes Us Healty?", New York Times Magazine, 16th September 2007. Available at: https://www.nytimes.com/2007/09/16/magazine/16epidemiology-t.html Imai Ch 2

Session 3: 17.09.2018 **Sampling and Distributions**	
Learning Objective	Probability and random variables; useful distributions; expected values; random sampling; and sampling distributions
Required Readings	Stock & Watson Ch 2
Optional Readings	Kellstedt & Whitten Chs 5--6 DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics.* Fourth Edition. Pearson Education. Chs. 1--3

Session 4: 24.09.2018 **T test for the Difference in Means and Hypothesis Testing**	
Learning Objective	Significance test; difference in means; fundamentals of all hypothesis testing
Required Readings	Stock & Watson Ch 3
Optional Readings	Kellstedt & Whitten Ch 7

Session 5: 01.10.2018 **T tests for the Difference in Means, Confidence Intervals and Revision – please bring substantial questions**	
Learning Objective	Finish t tests for the difference in means and confidence intervals – revision of material including sample variance; standard errors; t tests and confidence intervalse
Required Readings	Stock & Watson Ch 2 and Stock & Watson Ch 3
Optional Readings	

Session 6: 08.10.2018

Bivariate Linear Regression Models	
Learning Objective	Linear regression with one explanatory variable; comparing linear regression and t test for means; uncertainty; goodness of fit; prediction
Required Readings	Stock & Watson Ch 4 and 5.1--5.3
Optional Readings	Kellstedt & Whitten 8.1--8.4; Angrist & Pischke Ch 2

Session 7: 15.10.2018 **Multiple Linear Regression Models (I)**	
Learning Objective	Linear regression with multiple explanatory variables; control variables; F test for joint hypothesis testing
Required Readings	Stock & Watson Chs 6.2--6.4 and 7.1--7.3 and 7.5
Optional Readings	Kellstedt & Whitten Ch 9 James et al. Ch 3

Mid-term Exam Week: 22-26 October 2018 – no class

Session 8: 29.10.2018 **Multiple Linear Regression Models (II)**	
Learning Objective	Including qualitative information; interactions with dummies; interactions of continuous variables
Required Readings	Stock & Watson 8.3 and 8.4 Thomas Brambor, William R. Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analysis" *Political Analysis.* 14(1): pp. 63--82.
Optional Readings	Kellstedt & Whitten Ch 10.1--10.3

Session 9: 05.11.2018 **Regression Assumptions and Violations of Assumptions**	
Learning Objective	Omitted Variable Bias/ Conditional Mean Independence; Heteroskedasticity; Collinearity; Outliers
Required Readings	Stock & Watson Chs 5.4--5.6 and 6.1 and 6.5--6.7 and 9
Optional Readings	Kellstedt & Whitten Ch 8.5

Session 10: 12.11.2018 **Panel Data Structures: Fixed Effects Models**

Learning Objective	Cross sections and time series; unit fixed effects; time fixed effects; two-way fixed effects; heteroskedasticity; auto-correlation; spatial dependence
Required Readings	Stock & Watson Ch 10
Optional Readings	Kellstedt & Whitten Ch 11.3 Neal Beck and Jonathan Katz. 2011. "Modelling Dynamics in Time-Series—Cross-Section Political Economy Data" *Annual Review of Political Science.* 14: 331--352

Session 11: 19.11.2018 **Binary Outcomes: The Linear Probability Model**	
Learning Objective	Discrete and limited dependent variables; using the linear probability model; discussing other models for classification
Required Readings	Stock & Watson Ch 11.1
Optional Readings	Stock & Watson Ch 11.2--11.5 James et al. Ch 4 Kellstedt & Whitten Ch. 11.2

Session 12: 26.11.2018 **Putting It All Together or Randomized Controlled Trials (will be announced)**	
Learning Objective	Data analysis and evidence-based policy making and research
Required Readings	Nicholas Carners and Eric R. Hansen 2016: Does Paying Politicians More Promote Economic Diversity in Legislatures? *American Political Science Review.* Vol. 110(4), pp. 669-716.
Optional Readings	Kellstedt and Whitten Chs 1--4 and 12 King, Gary, Robert Keohane, and Sidney Verba. 1994. Designing Social Science Inquiry. Princeton: Princeton University Press. Gary King. 1995. "Replication, Replication." *Political Science and Politics.* 28(3): 444--452 Gary King. 2006. "Publication, Publication." *Political Science and Politics.* 39(1): 119--125.

Final Exam Week: 10-14 December 2018 – no class