POL40950 / POL41750
Introduction to Statistics

Johan A. Elkink

School of Politics and International Relations
University College Dublin

jos.elkink@ucd.ie
Newman Building, Rm F304
http://www.joselkink.net/teaching

Autumn 2016

Introduction

This course consists of two components, descriptive and inferential statistics. In the first half we will focus on descriptive statistics: How to plot numerical information in a useful way? How to look at the distribution of variables? How to generate numerical summaries of your data? How to look at correlations in your data between different variables? During the second half we turn to inferential statistics. Inference is drawing conclusions about facts you do not observe from facts you do. For example, in statistics it is common to draw a random sample of individuals from a population to gather data on those individuals in order to derive insights about the population as a whole. This part of the course will discuss the logic behind such inferences and some of the statistical tools available to do so. Since regression analysis, with its many extensions, is the most common approach for such inferences in political science, the course will focus on this technique, including briefly its extension to logistic regression – regression where the phenomenon to be explained is binary.

The class sessions will consists of a one hour lecture and a one hour lab session. The core textbooks for the course are Blalock (1979) and Moore, McCabe and Craig (2012), but most of the contents can be found in a large number of textbooks. Blalock (1979) is significantly more verbose and contains careful explanations of the topics, but is only limitedly available; Moore, McCabe and Craig (2012) is a more modern textbook, but at times rather brief in its descriptions.

Three further textbooks to highlight are Norris et al. (2012), which is very brief in its explanations, but provides extensive details on how to implement things in SPSS; Field (2009), which provides good explanations, but where the order in which topics are discussed is very different from this course – in particular, the distinction between descriptive and inferential statistics is not really present in this book; and Kellstedt and Whitten (2009), which is specific to political science and, after a more general introduction to research design, follows many classes in this course closely.

In addition, there will be handouts with a description of the software commands you will need to learn to be able to do the homeworks and apply the theoretical material in practice.

Grading

The only way to properly learn statistics is by hands-on training. You will need to work with actual data and produce your own statistical analysis – just the theory will never be sufficient. For that reason, there will be four practical homework assignments, each worth 25% of the grade. The assignments will be available online. The standard penalty for late submission as outlined in the SPIRe Masters Handbook will apply, whereby it should be taken into account that a late submission might result in a delayed return of feedback to the entire class.

Plagiarism

Although this should be obvious, plagiarism – copying someone else's text without acknowledgement or beyond "fair use" quantities – is not allowed. Please carefully check the UCD policies concerning plagiarism[1] and its more extensive description of what is plagiarism and what is not[2].

Classes

Classes take place once a week, with a lecture on Monday 1-2 PM in L143 in the Law Building and lab sessions Monday 2-3 PM in G5 of the Daedalus building, both at the Belfield campus of UCD.

---

[1]http://www.ucd.ie/regist/documents/plagiarism_policy_and_procedures.pdf.

[2]http://www.ucd.ie/library/students/information_skills/plagiari.html

Contact

I do not have fixed office hours, so if you want to make sure I am present, you can make an appointment by email. If a personal visit is not necessary, the easiest way to reach me is by email (jos.elkink@ucd.ie). The teaching assistant for this module will be Arya Pillai, who can be reached at arya.pillai@ucdconnect.ie.
Course materials will be uploaded to http://www.joselkink.net/teaching.

To stay up to date with developments in the UCD School of Politics and International Relations, also keep an eye on the following social media:
Web: http://www.ucd.ie/politics/
Blog: http://politicalscience.ie/
Twitter: http://twitter.com/ucdpolitics
Facebook: http://www.facebook.com/ucdspire

Homeworks

There will be four homework assignments during this module, each worth 25% of the grade. The first assignment will be due **Wednesday 5/10, 5 pm**, the second **Wednesday 26/10, 5 pm**, the third **Wednesday 16/11, 5 pm**, and the final assignment will be due **Wednesday 7/12, 5 pm**. Assignments should be submitted electronically to jos.elkink@ucd.ie, consisting of two files: a PDF file containing the written-up answers and a command file of the statistics package used (e.g. SPS file in SPSS, R file in R, or DO file in Stata). Note that Microsoft Word formats are *not* accepted and submitting in the wrong format first can cause your submission to be late. You are free to do the homeworks in SPSS, R, Stata, or any other package, as long as the appropriate command file is provided. Whichever software you use, command files need to be properly laid out, which includes:

- Include all commands that lead to the results used in the assignment, including opening files.

- Using only those commands that actually lead to the results used in the assignment, with no superfluous or additional commands.

- Using enough whitespace and indention to keep the code easy to follow.

- Inserting sufficient comments to clarify the commands.

The choice of software is up to you, but for PhD students or those who have a keen interest in developing strong statistical programming skills I would recommend R, for all others I would recommend Stata.

The grading will take into consideration the presentation of the results. Tables copy/pasted from SPSS or computer output with some interpretation squeezed in will get lower grades than assignments that are properly presented and formatted. When a substantive interpretation is asked for, this means an interpretation in terms of the substantive content of the topic at hand: e.g. if the data is on attitudes and turnout in elections, the substantive interpretation is "what do these results tell us about voting behaviour?", not "is this relationship positive or negative, significant or not?". The translation back from the statistical results to the political science interpretation is crucial to a good grade in this course.

12 September: Introduction: data & measurement

What is quantitative political science? What are data? What is a variable? What are the different levels of measurement?

required	Blalock (1979: ch 1-2)
alternatives	Pollock (2005: ch 1)
	Kellstedt and Whitten (2009: ch 5-6)
	Healey (2011: ch 1)
	Privitera (2011: ch 1)
	Argyrous (1997: ch 1, 5)
	Fielding and Gilbert (2000: ch 1)
SPSS	Field (2009: ch 3)
	Norris et al. (2012: ch 1)
	Hosker (2008: ch 1-2)
	Levesque (2004) (advanced)
R	Pollock (2014: ch 1, 3, 11)
	Field, Miles and Field (2012: ch 3)
	Verzani (2005: ch 1)
	Dalgaard (2002: ch 1)
	Maindonald and Braun (2007: ch 1)

19 September: Univariate descriptives

How to describe your variables graphically, including pie charts, histograms. How to describe your variables numerically, including the mean, mode, median, variance, and standard deviation.

required	Blalock (1979: ch 3-6)
	Moore, McCabe and Craig (2012: §1.1-1.2)
alternatives	Field (2009: §4.1-4.5)
	Privitera (2011: §2-4)
	Diamond and Jefferies (2001: ch 2-5)
	Healey (2011: ch 2-4)
	Wright and London (2009: ch 1-3)
	Argyrous (1997: ch 2-4)
	Kellstedt and Whitten (2009: ch 6)
	Heiman (2001: ch 6-8)
SPSS	Norris et al. (2012: ch 2-4)
	Fielding and Gilbert (2000: ch 3-5)
	Hosker (2008: ch 6, 8)
R	Pollock (2014: ch 2)
	Verzani (2005: §2.1-2.3)
	Dalgaard (2002: ch 3)
	Maindonald and Braun (2007: ch 2)

26 September: Multivariate descriptives

How to describe relations between variables graphically, including bar charts, scatter plots, box plots? Discussion of covariance and correlation to look at numerical indicators of relationships.

required	Blalock (1979: ch 16-17)
	Moore, McCabe and Craig (2012: §2.1-2.2, 2.5-2.6)
alternatives	Pollock (2005: ch 3-4)
	Privitera (2011: ch 15)
	Diamond and Jefferies (2001: ch 13)
	Healey (2011: ch 14)
	Heiman (2001: ch 10)
	Argyrous (1997: ch 22)
SPSS	Norris et al. (2012: ch 6-7)
	Fielding and Gilbert (2000: ch 6, 8-9)
R	Pollock (2014: ch 4)
	Field, Miles and Field (2012: ch 4)
	Verzani (2005: §3.1-3.3, ch 4)

3 October: Simple regression

Descriptive univariate linear regression models – how to look at the relation between two continuous variables?

required	Blalock (1979: ch 17)
	Moore, McCabe and Craig (2012: §2.3-2.4)
alternatives	Pollock (2005: ch 7)
	Privitera (2011: §16.1-16.5)
	Diamond and Jefferies (2001: ch 13)
	Healey (2011: ch 14)
	Heiman (2001: ch 11)
	Argyrous (1997: ch 22)
	Miles and Shevlin (2001: ch 1)
	Wright and London (2009: ch 8)
SPSS	Norris et al. (2012: ch 8)
	Fielding and Gilbert (2000: ch 8)
R	Pollock (2014: ch 8)
	Field, Miles and Field (2012: ch 6-7)
	Verzani (2005: §3.4)
	Dalgaard (2002: ch 5)
	Maindonald and Braun (2007: ch 5)
further	Starnes, Yates and Moore (2010: §2.6)

10 October: Probabilities and probability distributions

What are probabilities and probability distributions? Introduction to the normal distribution.

required	Hosker (2008: ch 3)
	Moore, McCabe and Craig (2012: §1.3)
alternatives	Privitera (2011: ch 5-6)
	Diamond and Jefferies (2001: ch 7)
	Healey (2011: ch 5)
	Argyrous (1997: ch 6)
	Fielding and Gilbert (2000: ch 7)
	Heiman (2001: ch 9)
SPSS	Norris et al. (2012: ch 5)
R	Dalgaard (2002: ch 2)
	Maindonald and Braun (2007: ch 3)

17 October: Sampling distributions and the Central Limit Theorem

What is statistical inference? Introduction to sampling methods. What is the Central Limit Theorem?

required	Kellstedt and Whitten (2009: ch 7)
	Blalock (1979: ch 7, 9)
	Moore, McCabe and Craig (2012: §3.2-3.3, §5.1)
alternatives	Pollock (2005: ch 5)
	Hosker (2008: ch 4)
	Privitera (2011: ch 7)
	Diamond and Jefferies (2001: ch 8)
	Healey (2011: ch 6)
	Heiman (2001: ch 12)
	Argyrous (1997: ch 7)
	Fielding and Gilbert (2000: ch 10)
	Wright and London (2009: ch 4)
further	Blalock (1979: ch 21)
	Moore, McCabe and Craig (2012: ch 3-4)
R	Verzani (2005: ch 5)
	Maindonald and Braun (2007: ch 4)

24 October: Hypothesis tests and confidence intervals

What are hypothesis tests and how do I generate confidence intervals?

required	Blalock (1979: ch 8, 12)
	Moore, McCabe and Craig (2012: ch 6)
	Kellstedt and Whitten (2009: ch 8)
recommended	Senn (2012)
alternatives	Pollock (2005: ch 6)
	Hosker (2008: ch 5, 9-)
	Privitera (2011: ch 8, 11)
	Diamond and Jefferies (2001: ch 9-11)
	Healey (2011: ch 7)
	Heiman (2001: ch 13-14)
	Argyrous (1997: ch 8-9)
	Fielding and Gilbert (2000: ch 11)
	Wright and London (2009: ch 5)
SPSS	Norris et al. (2012: ch 10-12)
R	Pollock (2014: ch 6-7)
	Verzani (2005: ch 7)

7 November: Comparing means and proportions

How to compare the means or proportions of two different groups?

required	Blalock (1979: ch 11, 13)
	Moore, McCabe and Craig (2012: ch 7-8)
	Kellstedt and Whitten (2009: ch 8)
alternatives	Field (2009: ch 9)
	Privitera (2011: ch 9-10)
	Diamond and Jefferies (2001: ch 12, (14))
	Healey (2011: ch 8-9)
	Argyrous (1997: ch 10-11, 13, 17)
	Wright and London (2009: ch 6)
SPSS	Norris et al. (2012: ch 10-11)
R	Field, Miles and Field (2012: ch 9)
	Verzani (2005: ch 8)
	Dalgaard (2002: ch 4)
further	Moore, McCabe and Craig (2012: ch 9)

14 November: Multiple regression: t-tests and F-tests

How to test for significance in regression models? If I see a correlation between two variables, how do I know whether this is due to chance or "real"? How to perform and interpret regression models with more than one independent variable?

required	Kellstedt and Whitten (2009: ch 9-10)
	Blalock (1979: ch 16, 18-19)
	Moore, McCabe and Craig (2012: ch 10-11)
	Kellstedt and Whitten (2009: ch 10)
recommended	Lewis-Beck (1980)
alternatives	Privitera (2011: ch 16)
	Argyrous (1997: ch 22)
	Hosker (2008: ch 10-11)
	Field (2009: ch 7)
	Healey (2011: ch 16)
	Miles and Shevlin (2001: ch 2)
SPSS	Norris et al. (2012: ch 13-14, 19)
R	Pollock (2014: ch 5, 8)
	Field, Miles and Field (2012: ch 7)
	Verzani (2005: ch 8, §10.3, ch 11)
	Dalgaard (2002: ch 9)
	Maindonald and Braun (2007: ch 6)
further	Moore, McCabe and Craig (2012: ch 9)
	Miles and Shevlin (2001: ch 4-5)
	Allison (1999)

21 November: Multiple regression – categorical independent variables

Categorical independent variables and simple interactions in multiple regression.

required	Kellstedt and Whitten (2009: ch 11-12)
	Blalock (1979: ch 20)
recommended	Hardy (1993)
alternatives	Miles and Shevlin (2001: ch 3)
R	Pollock (2014: ch 9)

28 November: Multiple regression – categorical dependent variables

Regression analysis when the dependent variable is binary – e.g. explaining turnout in elections. Introduction to logistic regression.

required	Moore, McCabe and Craig (2012: ch 14)
recommended	Pampel (2000)
alternatives	Pollock (2005: ch 8)
	Menard (2002)
R	Pollock (2014: ch 10)
	Field, Miles and Field (2012: ch 8)
	Dalgaard (2002: ch 11)
	Maindonald and Braun (2007: ch 8)

References

Allison, Paul D. 1999. *Multiple regression: a primer*. Thousand Oaks, CA: Pine Forge Press.

Argyrous, George. 1997. *Statistics for social research*. Basingstoke: MacMillan.

Blalock, Hubert M. 1979. *Social statistics*. 2nd ed. Tokyo: McGraw-Hill Kogakusha.

Dalgaard, Peter. 2002. *Introductory statistics with R*. New York, NY: Springer.

Diamond, Ian and Julie Jefferies. 2001. *Beginning statistics: an introduction for social scientists*. London: Sage.

Field, Andy. 2009. *Discovering statistics using SPSS*. 3rd ed. London: Sage.

Field, Andy, Jeremy Miles and Zoë Field. 2012. *Discovering statistics using R*. London: Sage.

Fielding, Jane and Nigel Gilbert. 2000. *Understanding social statistics*. London: Sage Publications.

Hardy, Melissa A. 1993. *Regression with dummy variables*. London: Sage.

Healey, Joseph F. 2011. *Statistics: a tool for social research*. 9th ed. Belmont CA: Wadsworth.

Heiman, Gary W. 2001. *Understanding research methods and statistics: an integrated introduction for psychology*. 2nd ed. Boston: Houghton Mifflin.

Hosker, Ian. 2008. *Starting statistics: Data handling for beginners*. Abergele: Studymates.

Kellstedt, Paul M. and Guy D. Whitten. 2009. *The fundamentals of political science research*. Cambridge: Cambridge University Press.

Levesque, Raynald. 2004. *Programming and data management for PASW Statistics 18. A guide for PASW Statistics and SAS users*. Chicago: SPSS Inc.

Lewis-Beck, Michael S. 1980. *Applied regression: an introduction*. Sage Publications.

Maindonald, John and John Braun. 2007. *Data analysis and graphics using R. An example-based approach*. 2nd ed. Cambridge: Cambridge University Press.

Menard, Scott. 2002. *Applied logistic regression analysis*. 2nd ed. London: Sage.

Miles, Jeremy and Mark Shevlin. 2001. *Applying regression & correlation: a guide for students and researchers*. London: Sage Publications.

Moore, David S., George P. McCabe and Bruce A. Craig. 2012. *Introduction to the practice of statistics*. 7th international edition ed. New York: W.H. Freeman.

Norris, Gareth, Faiza Qureshi, Dennis Howitt and Duncan Cramer. 2012. *Introduction to statistics with SPSS for social science*. Harlow: Pearson.

Pampel, Fred C. 2000. *Logistic regression: A primer*. London: Sage.

Pollock, Philip H. 2005. *The essentials of political analysis*. 2nd ed. Washington, DC: CQ Press.

Pollock, Philip H. 2014. *An R companion to political analysis*. Thousand Oaks, CA: CQ Press.

Privitera, Gregory J. 2011. *Statistics for the Behavioral Sciences*. Sage.

Senn, Stephen. 2012. "Tea for three. Of infusions and inferences and milk in first." *Significance* (12):30–33.

Starnes, Daren S., Dan Yates and David S. Moore. 2010. *The practice of statistics*. 4th ed. New York: W.H. Freeman.

Verzani, John. 2005. *Using R for introductory statistics*. Boca Raton, FL: Chapman & Hall/CRC.

Wright, Daniel B. and Kamala London. 2009. *First (and second) steps in statistics*. 2nd ed. Los Angeles: Sage Publications.