Data Analysis in Social Science 3 (SOC2094/3094 / POL2094/3094)

Dr Alexey Bessudnov

December 2016

1 Practical arrangements

Lectures:

- Tuesday, 4.30 - 5.30 pm, WSL 220

Lectures start on 24 January and there will be six lectures all together. The last lecture will be on 28 February.

Computer lab sessions:

- Thursday, 10.30 am - 12 pm, Forum Exploration Lab 1

The computer lab sessions start on 26 January. There will be an additional computer lab session on Tuesday 7 February 9.30 to 11 am in Forum Exploration Lab 1.

Office hours (Amory A341):

- Tuesday, 11am - 12pm

- Thursday, 12pm-1pm

Email:

- A.Bessudnov@exeter.ac.uk

2 Aims of the module

This is a third module in the series of modules on data analysis in the social sciences. In Data Analysis 1 you studied the basics of descriptive statistics and statistical inference. Data Analysis 2 mostly covered linear regression analysis. In Data Analysis 3 we are not going to learn new statistical techniques, but will learn how to apply the techniques you already know to the analysis of real-life data sets and how to produce good statistical reports.

This is a skill that you may need in a variety of jobs where data analytic expertise is required, such as marketing analysis, policy analysis in various fields, internet analytics, data journalism, academic research, etc.

You already now how to use R to describe data and run simple statistical models. However, real-life data rarely come in the form of a perfectly formatted csv file ready for the analysis. The real life data sets often need to be reshaped, merged, recoded, aggregated and modified in various ways before you can even start your analysis. Unless you know how to do this you will not be able to produce good statistical reports. Apart from these techniques we will learn how to effectively communicate the results of your statistical analysis.

The data for this module come from the University of Exeter's Planning department. These are data about the modules students take in Exeter and the socio-demographics of students. These are complex data sets and we will work together with Planning to make sense of these data that may help the university with its educational policies.

Throughout the module we will use R for statistical analysis. You are expected to know the basics of data analysis in R.

The pre-requisites for this module are POL/SOC1041 and POL/SOC2077.

3 Attendance

This module is very technical. As with other technical skills missing some initial bits means that you will not be able to make it. Attendance in this module is crucial. If you do not attend you will not be able to do well in this module. Even skipping a couple of classes may have very negative consequences for how you understand the material. Another negative consequence will be that you will slow the rest of the class down as I will have to explain the same things several times. If you plan not to attend the classes please do not take this optional module.

4   Assessment

The assessment for this module is a report of 4,000 words (in addition to figures and tables) with the results of statistical analysis you will undertake. This will be 100% of your final mark for this module. You will be given questions for the reports later in the module. The analysis will use the data about students in the University of Exeter.

The deadline for submitting your reports through eBart is 4 April at 2pm. You will receive your marks and feedback by 5 May.

Late submissions up to two weeks after the deadline will be capped at 40%. Submissions that are late for more than two weeks will not be accepted.

5   Syllabus plan

This will be the first delivery of this module and I may change some topics as we proceed.

- Data structures in R (another R refresher at a more advanced level)

- Manipulating data with dplyr

- Longitudinal data in R. Wide and long formats. Reshaping

- Data visualisation with ggplot2

- Producing statistical reports with R Markdown

- Interactive applications with Shiny

- Loops and other control structures. The apply family of functions

- Writing functions in R

At the computer lab sessions we will be working with the real-life data applying the techniques you learn at the lectures.

6   Reading list

I will recommend the readings/chapters for each lecture / computer lab as the course progresses.

- G.Grolemund & H.Wickham. (2016). R for Data Science. Freely available at `http://r4ds.had.co.nz/`.

- H.Wickham. (2015). ggplot2. Elegant Graphics for Data Analysis. 2nd ed. Springer.

- W.Chang. (2013). R Graphics Cookbook. O'Reilly.

- P.Spector. (2008). Data Manipulation with R. Springer.

- N.Matloff. (2011). The Art of R Programming. No Starch Press.

- H.Wickham. (2014). Advanced R. Chapman & Hall.