LEAPS

Summer School

2017

An introduction to data analysis

Course Co-ordinator: Dr Kevin Ralston



An introduction to data analysis

**Introduction**

Processing, managing and analysing data is important in many fields of employment and education. This course is relevant to all students going to university, but it is of particular interest to those who will go into subjects involving statistical analysis, such as business and/or social science. Students will have the chance to familiarize themselves with principles of data analysis at a first year undergraduate level. Important parts of the research process will be introduced, including: data sources, analysis packages, formulating questions, descriptive statistics, and programming. In addition, a significant proportion of course time will be spent in a PC lab engaging with data and learning about practical aspects of data analysis. The programme of work is quite flexible and will address the needs and progress of each individual student. Assessment for this course is based on coursework assignments and an end of course test. As a data analyst taking the course you will be encouraged to undertake a short piece of analysis on an area of interest to you.

Data analysis is fun (really) and good data analysts apply many kinds of skills in the work they do! This includes maths, computer programming and communication skills.

**Learning outcomes and course aims[[1]](#footnote-1)**

* A learner should be able to apply data analysis techniques
* A data analysis should be able to use good principles of ‘data management’
* Data analysts will be able to interpret statistical tests and computer outputs
* A researcher should have the ability to undertake a piece reproducible data analysis
* A data analyst should be able to use appropriate software

The course will provide data analysts with the resources to achieve these outcomes. Learners will be introduced to this at an introductory level. No previous experience is assumed. The emphasis is placed on actively building experience. This course will provide a basis of understanding and resources that can be used for future data analysis a learner may undertake.

**Content**

Statistical methods

Confidence intervals, statistical significance, contingency tables, association, regression, probability distributions

Computer software, e.g. SPSS, Excel

Programming syntax, reading data in, saving data out, employing statistical techniques, graphing and summarizing data

Data management

Cleaning, recoding, organising, peer programming, creating reproducible annotated files

**Learning approaches**

Data analysts taking the course will actively engage with secondary data sources and also generate data in class. This will enable data analysts to practice using statistical methods. Readings on concepts and methods will be provided. Data analysts will also be introduced to online resources such as visual simulations and talks. Principles, such as peer programming and annotating syntax will be employed.

|  |  |  |
| --- | --- | --- |
| 1,Tues 6th June | Introduction  A look at the ‘research process’ and starting with the software (SPSS). Syntax and read data into the package. |  |
| 2, Fri 9th June | Data, what are data and where to find it! |  |
| 3, Tues, 13th  June | Exploring data with graphs and numerical summaries |  |
| 4, Fri  16th June | Association: contingency, correlation, and regression |  |
| 5, Tues  20th June | Gathering data |  |
| 6, Fri 23rd June | Probability distributions (normal) |  |
| 7, Tues  27th June | Sampling distributions | Hand in course work 1 |
| 8, Fri  30th June | Statistical inference: confidence intervals |  |
| 9, Tues  4th July | Statistical inference: significance tests |  |
| 10, Fri  7th July | Project work |  |
| 11, Tues  11th July | Comparing two groups |  |
| 12, Fri  14th July | Project work |  |
| 13, Tues  18th July | Categorical association and regression | Hand in course work 2 |
| 14, Fri  21st July | Class test |  |

**Assessment**

30% Coursework, write a research proposal

40% Coursework, write a short piece of replicable analysis

30% Class test

**Reading Material**

Students will be guided to any reading material necessary to complete the course. Many of these resources are available online.

Here is a free book which introduces data management in SPSS:

<http://www.spsstools.net/en/spss-programming-book/>

Here is a great free text-book introduction to statistics, take a look:

<https://d3bxy9euw4e147.cloudfront.net/oscms-prodcms/media/documents/Introductory_Statistics-OP.pdf>

Here is another one:

<http://onlinestatbook.com/Online_Statistics_Education.pdf>

Here is a book, from the American Statistical Association, which introduces surveys and sampling:

<https://www.whatisasurvey.info/overview.htm>

**Class times and locations**

Data analysis will be on Tuesday mornings (9:30-12:15) and on Friday afternoons (13:45-16:30).

The location will be room g.06 at Thomson's Land in Moray House. The first class will start promptly at 9:30 on Tuesday 6th June.

1. The course content and principles introduced are applicable in many disciplines [↑](#footnote-ref-1)