Data Analysis

We asked our university partners for their advice on carrying out data analysis. This guide summarises the approaches they recommend and offers suggestions on different pieces of software and programming languages to use.

1. Keep it simple

When setting your research aim(s) or formulating a question you are trying to answer keep it as simple as possible.

"If you can state your question clearly and simply your analysis is likely to be easier to conduct", Professor Andy Foggo – University of Plymouth

2. To plan or not to plan?

Research scientists take different approaches to analysis. Some will make a plan before they collect the data whilst others don't.

Huw Morgan of the University of Aberystwyth takes the latter approach: *"Play with the data. Do not approach the data set with a rigid analysis plan. Spend time on the analysis experimenting with different approaches, testing various methods."*

However, something that all research scientists agree on is that it's preferable to have too much data than to have too little: *"remember that the power of data to allow us to arrive at a confident answer to a question is strongly affected by the number of data we have; I've never had a problem as a result of having too many data but I've certainly been frustrated by having too few."* – Professor Andy Foggo.

3. Getting started

Some general advice is to start making plots of the different variables in order to look for a correlation. However it is also important to be flexible: Dr Robert Beardmore of the University of Exeter recommends that you should *"never be fixed to one methodology, keep an open mind and be prepared to learn new techniques when the ones you know don't work".*

Professor Andy Foggo recommends some approaches: *"Don't underestimate the power of creating simple summary graphics: a bar chart or a box and whisker plot showing averages and a measure of variability (mean plus or minus standard deviation would be a good example) is a great form of graphical analysis when comparing groups, a scatterplot does the same for relationships between variables"*.

Data Analysis

4. Does it make sense?

It is important to ensure that the data you have collected or been given does not contain errors. These errors could be down to faults with the equipment or mistakes when it has been recorded.

If you are using data that will have been looked at by other people, Dr Peter Hatfield from the University of Oxford recommends you *“look at other things people have done with the data, so (a) you don’t waste time duplicating efforts and (b) can check you are getting sensible results”*.

You can also compare your data with different but complementary data sets to see if you are getting similar results. For example, if you are carrying out an experiment using the MX-10 radiation detector, you can search TAPAS to find data taken by another school who have carried out the same experiment.

5. Ask for advice!

Research scientists will work together to solve problems. It is very common for them to approach others to ask for advice, you should do the same. You can either research and find a contact and approach them yourself or contact us and we can help put you in touch with someone who can help.

Data Analysis

Useful Tools for Analysis

Research scientists use a broad range of packages to examine data. Some suggestions are made below but this list is not exhaustive. Many will write their own analysis packages using a programming language. This has the benefit of being exactly tailored to your data. However, general packages will have built-in tools for lots of different types of analysis so there are benefits to both.

- Tableau for data analysis and visualisation (free to IRIS schools, contact us for details).
- MS Excel for data management, manipulation and storage
- PRIMER and PERMANOVA are used for specialist analyses in ecology
- ImageJ for quick analyses of images
- Topcat for looking at astronomical data: http://www.star.bris.ac.uk/~mbt/topcat/

Coding-related:

- C
- Github (platform for sharing and developing code https://github.com/)
- Java
- Matlab
- Minitab
- Python
  - Anaconda: https://www.continuum.io/why-anaconda
  - Code Academy is also a way to get started: https://www.codecademy.com/learn/python
- R and R Studio
- SPSS

File Types

You will come across multiple filetypes in the process of your research. These will range from picture files (.png or .jpeg) to documents holding large amounts of numerical information (.CSV). There are those which are specific to the discipline, for example astronomical data can often be presented in .FITS files.

Each of the different types will need to be opened with the appropriate software package or file manager. A comprehensive list of file types can be found here: https://en.wikipedia.org/wiki/List_of_file_formats

Data Analysis

Thanks

We are very grateful to the advice shared by the following researchers (they are listed along with their research interests):

- Professor Robert Beardmore, University of Exeter

  Mathematics (geometric problems in singular dynamics systems), evolutionary microbiology, mathematical modelling of evolutionary processes, DNA analysis of evolved microbes, antibiotic resistance evolution.

- Professor Andy Foggo, University of Plymouth

  Ecology: working with marine ecosystems and organisms

- Dr Peter Hatfield, University of Oxford
  Astrophysics
- Dr Huw Morgan, University of Aberystwyth

  Astronomy: the Sun's atmosphere