APS 240: Data Analysis and Statistics with R

Dylan Z. Childs

2016-10-11

Contents

Chapter 1

Course information and overview

This is the online course book for the **Data Analysis and Statistics with R** (APS 240) module. You can view this book in any modern desktop browser, as well as on your phone or tablet device. The site is self-contained—it contains all the material you are expected to learn this year.

Dylan Childs is the course co-coordinator. Please email him if you have have any general queries about the course. Andrew Beckerman is the second course instructor. The Teaching Assistants ('TAs') this year are Ross Booton, Matthew Hethcoat, Bethan Hindle, Tamora James, Felix Lim, and Simon Mills.

1.1 Why do a data analysis course?

To do science yourself, or to understand the science other people do, you need some understanding of the principles of experimental design, data collection, data presentation and data analysis. That doesn't mean becoming a maths wizard, or a computer genius. It means knowing how to take sensible decisions about designing studies and collecting data, and then being able to interpret those data correctly. Sometimes the methods required are extremely simple, sometimes more complex. You aren't expected to get to grips with all of them, but what we hope to do in the course is to give you a good practical grasp of the core techniques that are widely used in biology and environmental sciences. You should then be equipped to use these techniques intelligently and, equally importantly, know when they are not appropriate, and when you need to seek help to find the correct way to design or analyse your study.

You should, with some work on your part, acquire a set of skills which you will use at various stages throughout the remainder of your course, in practicals, field courses and in your project work. These same skills will almost certainly also be useful after your degree, whether doing biology, or something completely different. We live in a world that is increasingly flooded with data, and people who know how to make sense of this are in high demand. The R statistical programming environment underpins much of this endeavour, in both academic and commercial settings. Learning the basic principles of data analysis with R will only improve your employment prospects.

1.2 Course overview

1.2.1 Aims

This course has two main, and equal, aims. The first is to provide a basic training in the use of statistical methods and software (R and RStudio) to analyse biological data. The second is to introduce some of the principles of experimental design, sampling, data interpretation, graphical presentation and scientific writing relevant to the biological and environmental sciences.