BGY2011M Data Skills for the Life Sciences

Module lead: IAIN STOTT / email istott@lincoln.ac.uk

Last updated 2020-08-20

Contents

Chapter 1

Introduction

Welcome to Data Skills for the Life Sciences! This page contains important information on the module: it's worth reading through carefully, and the information will always be here for you to refer back to. The first few sections describe everything you need to know to set yourself up for the course. It's really important you complete these as soon as possible. Information on the timetable and learning schedule, and both formative and summative assessments, are found at the bottom of the page.

1.0.1 The analytical process

This module is all about analysing data. Any analytical process is really just a means of understanding your data, and relationships between variables in that data. *That's why I do not consider this to be a course in statistics; it is a course in giving you the skills to understand data.* Statistics is one part of that.

It's important not to dive right into statistical analysis without understanding your data, your variables and the relationships between them first. For that reason, much of the course will be dedicated to understanding data, and how to plot data.
These steps forms the **FIRST HALF of the course**:

- **Tidy data**
  - data exploration
  - data subsetting
  - data manipulation
- **Plots**
  - Single-variable distributions and summary statistics
  - Two- and three-variable plots to explore hypotheses

Data handling and plotting is also the best vehicle to learn how to code: data handling and plots are harder to code but easier to interpret, whereas statistical analyses are easier to code but harder to interpret.

Equally, you cannot complete a statistical analysis without assessing how well that model fits the data, and whether any 'assumptions' of your statistical model are violated.
These steps form the **SECOND HALF of the course**:

- **Model**
  - fitting a statistical model
- **Model fit**
  - assessing how well a statistical model fits the data
- **Results**
  - Understanding and interpreting the results of a statistical model

The "statistical model" we will be working with is the *general linear model (glm)*. This is a framework which brings together many traditional statistical analyses. You will learn where these traditional analyses fit into the glm framework.

There are naturally feedbacks between these steps, where new information may mean you need to revise earlier steps.

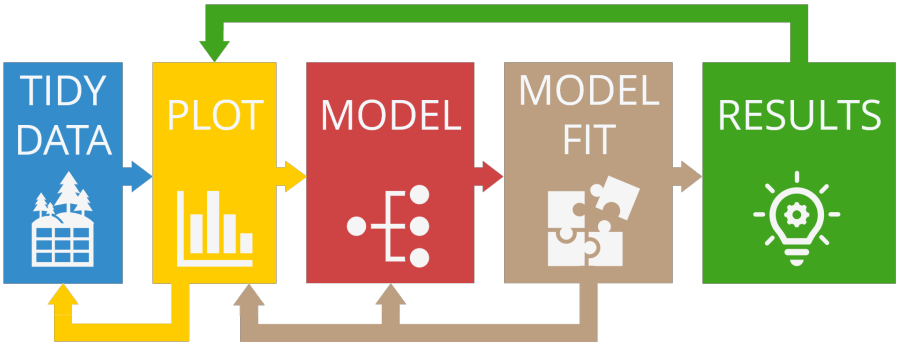

Figure 1.1: The analytical process and its feedbacks, from start to finish

1.1 Installing R and RStudio

Install here

Installing R and RStudio video: https://youtu.be/d7lARa168gk