

CPACT Data Analysis, Statistical Methods and R

Dr. Lee Fawcett



Introduction

Books and other resources

Some useful books include:

- W.N. Venables and D.M. Smith: An Introduction To R. Free to download at:
<https://cran.r-project.org/doc/manuals/R-intro.pdf>
- D. Braun and D.J. Murdoch: A First Course In Statistical Programming With R
- S. Ross: A First Course in Probability
- N.A. Weiss, P.T. Holmes and H. Michael: A Course In Probability

General background

Today we will learn about computational tools for performing probability calculations and statistical analysis.

We will use the R programming language.

R is very widely used in universities, businesses and other organizations due to its many advantages over other similar software packages.

- It is available for **free**
- It is '**open source**' and can be used for a very wide range of statistical and scientific applications
- **Libraries** of R functions and data-structures have been created and are readily available on the web

General background

Learning to use R is important because:

- (i) It is the foremost tool in **modern computational statistics**
- (ii) It is a great way to learn general concepts in **programming** which can be used with other languages or in other contexts
- (iii) It can be used to illustrate ideas in **probability and statistics**

Preliminaries

R is installed on all University machines; it *should* be installed on the machine you're sitting at now!

Today we will actually be using **RStudio**, which is an R *Integrated Development Environment*, or IDE.

This just makes R much easier to use and navigate, through the four windows that you see when you open RStudio (text editor, console window, file explorer and graphics display).

In this brief intro, we will quickly review:

- (i) Data types and vectors in R
- (ii) Dataframe manipulation
- (iii) Data summaries
- (iv) Functions and control statements

Please go to:

https://mas-coursebuild.ncl.ac.uk/lti/content/mathsaid-r/2021-22/?auth_level=0