A First Course In Statistics

Andrea Mascaretti

2020-02-06

Contents

1 Prerequisites **5**
- 1.1 Getting started . . . 5
- 1.2 Some useful links . . . 6
- 1.3 Licence . . . 6

2 Introduction **7**
- 2.1 Installing R . . . 7
- 2.2 Installing RStudio . . . 7
- 2.3 Installing packages . . . 7
- 2.4 Installing **moxier** . . . 8
- 2.5 Using **moxier** . . . 8

3 Introduction to programming in R **11**

4 Data Visualisation **13**
- 4.1 Base **R** . . . 13
- 4.2 **ggplot2** . . . 13

5 Descriptive Statistics **15**
- 5.1 Univariate Statistics . . . 15
- 5.2 Multiavriate Statistics . . . 15

6 Dimensionality Reduction **17**
- 6.1 Example 1 . . . 17
- 6.2 Example 2 . . . 17
- 6.3 Exercise 1 . . . 17

Chapter 1

Prerequisites

Welcome to this introductory course in statistics with R! This book will help you get started and will guide you through the material of the course.

The material of this course is based on the **moxier** package. It consists of a series of **learnr** notebooks, (Schloerke et al., 2019), that will guide through your first steps into the wonderful world of statistics and statistical computing!

1.1 Getting started

Whilst we will dedicate Chapter 2 to the topic of software installation, we here introduce the main tools we will use throughout this course: **R** and **RStudio**.

1.1.1 R

So, what do we talk about when we talk about R? According to the R project website, R is a software environment that includes:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

Quite a number of things! We will use this set of tools to dive deep into the principles of statistics and statistical computation.

Another thing that is worth noticing is that **R** is Free Software. It means anybody can contribute to its development. Common tools have emerged to solve problems. **R** can be extended with such tools, which are called *packages* in R parlance, to do all sort of incredible things. Packages are usually stored on CRAN, the Comprehensive R Archive Network, and they range from packages to send emails to machine learning.

1.1.2 RStudio

RStudio is an **R** IDE (Integrated Development Environment). But what does it mean? Basically, it is a set of software that helps you be more productive: it allows you to quickly manage files, see what variables you have defined and a vast number of other things. This book and the **moxier** package have been developed from within **RStudio**. We will see in a minute how to install it!

1.2 Some useful links

If you are interesting in learning more about **R** as a programming language, you can find many resources on the Internet. Some nice books are **Hands-On Programming with R**, (Grolemund, 2014) to get started and **Advanced R** (Wickham, 2019) to dive deep into the features of the language.

1.3 Licence

The **moxier** package is subject to the GPL-3 licence. For more information, visit https://mascaretti.github.io/moxier/.

Chapter 2

Introduction

We are now going to install the tools we need. We will start from **R**, to then move to **RStudio** and, finally, the packages we are going to need.

2.1 Installing R

Installing **R** has never been easier! You can simply follow the instructions you find in the video.

install-R from RStudio, Inc. on Vimeo.

An alternative is Microsoft R Open, which can be downloaded here. It lags behind the current **R** version, but it comes with better performance tweaks included.

2.2 Installing RStudio

Once you have got **R** installing, simply follow this other video.

Install RStudio from RStudio, Inc. on Vimeo.

2.3 Installing packages

We now cover the installation of packages. Packages extend the amount of things you can do with **R**, from plotting to data analysis. We will be using the **moxier** package through this course, for instance.

To quickly get up to speed as to how install packages, again, simply follow the video. Just, for the time being, skip the **tidyverse** installation.

install Packages from RStudio, Inc. on Vimeo.

2.4 Installing moxier

We are now ready to install **moxier**! First, we need to install an auxiliary package, the **remotes** package. This package allows to install other packages from GitHub. GitHub is a place where many people develop and publish open-source software.

To install **remotes**, simply type in your console

```
install.packages("remotes")
```

Once the installation is complete, you can install **moxier**

```
remotes::install_github("mascaretti/moxier")
```

2.5 Using moxier

We are now interested in using **moxier**. What is inside it? To find out, we list the tutorials contained in the package.

Name	Title
00-introduction-installing-r	R and RStudio
01-introduction-programming-r	Programming in R
02-data-visualisation-graphics	Graphics in Base R
03-data-visualisation-ggplot2	Introduction to ggplot
04-descriptive-statistics-univariate	Univariate Descriptive
05-descriptive-statistics-depth-measures	Depth Measures
06-descriptive-statistics-multivariate	Univariate Descriptive
07-descriptive-statistics-curse-dimensionality	Univariate Descriptive
08-dimensionality-reduction-pca	PCA: Example 1
09-dimensionality-reduction-pca	PCA: Example 2
10-dimensionality-reduction-pca	PCA: Exercise 1
11-hypotheses-test-anova	HT and ANOVA: Exercise 1
12-hypotheses-test-anova	HT and ANOVA: Exercise 2
13-linear-regression	Regression: Exercise 1
14-linear-regression	Regression: Exercise 1
15-classification	Classification - Cytokine
16-classification	Classification - Human Data
17-clustering	Clustering - 00
18-clustering	Clustering - Iris
19-clustering	Clustering - Earthquakes
20-extra-dplyr-intro	dplyr: An Introduction

Now, let us imagine you wish to run the tutorial on how to install **R**.

Simply run

```
learnr::run_tutorial(name = "00-introduction-installing-r", package = "moxier")
```

Quite smooth, right? Of course, put the title you are interested in to run the respective tutorial. To see the full list,

```
learnr::available_tutorials(package = "moxier")
```

Voilà! You are ready to start learning!

Chapter 3

Introduction to programming in R

This tutorial will introduce you to programming in **R**. Run it by doing

```
learnr::run_tutorial(name = "01-introduction-programming-r",
                     package = "moxier")
```

Chapter 4

Data Visualisation

In this part of the course, we will dive deep into data visualisation. We will start using base **R** and then explore the package **ggplot2**, (Wickham et al., 2019).

4.1 Base R

```
learnr::run_tutorial(name = "02-data-visualisation-graphics",
                     package = "moxier")
```

4.2 ggplot2

```
learnr::run_tutorial(name = "03-data-visualisation-ggplot2",
                     package = "moxier")
```

Chapter 5

Descriptive Statistics

We now turn our attention to descriptive statistics.

5.1 Univariate Statistics

5.1.1 Descriptive Univariate

```
learnr::run_tutorial(name = "04-descriptive-statistics-univariate",
                     package = "moxier")
```

5.1.2 Depth Measures

```
learnr::run_tutorial(name = "05-descriptive-statistics-depth-measures",
                     package = "moxier")
```

5.2 Multiavriate Statistics

5.2.1 Descriptive Multivariate

```
learnr::run_tutorial(name = "06-descriptive-statistics-multivariate",
                     package = "moxier")
```

5.2.2 Curse of Dimensionality

```
learnr::run_tutorial(name = "07-descriptive-statistics-curse-dimensionality",
                     package = "moxier")
```

Chapter 6

Dimensionality Reduction

We will study Principal Component Analysis.

6.1 Example 1

```
learnr::run_tutorial(name = "08-dimensionality-reduction-pca",
                     package = "moxier")
```

6.2 Example 2

```
learnr::run_tutorial(name = "09-dimensionality-reduction-pca",
                     package = "moxier")
```

6.3 Exercise 1

```
learnr::run_tutorial(name = "10-dimensionality-reduction-pca",
                     package = "moxier")
```

Bibliography

Grolemund, G. (2014). *Hands-On Programming with R.* O'Reilly. ISBN 978-1449359010.

Schloerke, B., Allaire, J., and Borges, B. (2019). *learnr: Interactive Tutorials for R.* R package version 0.10.0.

Wickham, H. (2019). *Advance R.* Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-0815384571.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., and Yutani, H. (2019). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* R package version 3.2.1.