Writing a book with R

An introduction

COLIN BOUSIGE[1]
Laboratoire des Multimatériaux et Interfaces, Lyon, France

This is a template for writing a book using the LaTeX template in `book_template.tex`. Render it by running:

```
bookdown::render_book("index.Rmd", "bookdown::pdf_book")
```

Version date: 07/04/2022
[1] Email: colin.bousige@univ-lyon1.fr

Chapter 1

Welcome

1.1 This is a section name

Hello ! [1, 2]
This is a nice template, don't you think ??

Chapter 2

About the class

2.1 Objectives of the class

The goal of this class is that at the end, the students are able to:

- Treat their data with the **free** and **open source** language **R**, *i.e.*:
  - Read, browse, manipulate and plot their data
  - Model or simulate their data
- Make automatic reporting through **Rmarkdown**
- Build a graphical interface with **Shiny** to interact with their data and output something (a value, a pdf report, a graph...)

What you will learn here will be useful in **any** scientific domain. The examples in this course are however mainly coming from the type of data you might encounter in Materials Science because, well, it's what I have on hand...

Chapter 3

A little reminder on Statistics

3.1 Why are statistical tools necessary in physical science?

When doing Science, one has to fully grasp the concept of *physical measurement*. Let's take an example to visualize the importance of this concept.

3.1.1 A practical example

Let's say you want to communicate to someone a temperature, and tell this person that the temperature is "38". If this is a random person in the street, they might think: "nice, let's go to the beach today!". If this random person is from the USA, they're gonna think: "damn, where did I put my coat?". If that person happens to be a physician, they might think: "that kid's got a slight fever". If they are a physicist doing a cryostat experiment, they might think "let's check the He tank level"... you see that one of the most important part of the measurement is missing: its unit. Units are there so that people understand each other when exchanging data, and you see here that 38 Celsius, 38 Fahrenheit or 38 Kelvin are quite different, and this quantity will mean different things in different contexts. A physical quantity given without its unit would be absolutely meaningless (unless, of course, you are looking at a unit-less quantity, like a count).

Now let's consider the body temperature of 38 °C given to a physician. How did you measure this temperature? With a mercury graduated thermometer or with a thermocouple? In the first case, you can probably assume that this value is given with a measurement error of at least 1 °C, meaning that the temperature you give to the physician is (38±1) °C, *i.e.* the physician won't be able to decide whether they should be concerned or not. In the second case, the temperature is often given with a 0.1 °C precision, so the physician, seeing that the body temperature is (38±0.1) °C, will probably tell you to take an aspirin and rest instead of giving you something stronger to treat a possible infection. Given that the uncertainty on the given value is of 0.1 °C, one should in fact give the temperature with matching decimal precision, *i.e.* (38.0±0.1) °C. Writing (38±0.1) °C, (38.00001±0.1) °C or (38.00±0.10000) °C would be meaningless too.

With this, we see that a physical measurement should be given with four parts: its actual **value**, its **decimal precision**, its **uncertainty**, and its **unit**. Should any of these four parts be missing in a physical quantity that you wanted to share, it would at best be imprecise, and at worst be utterly meaningless.

3.1.2 Probabilistic description of physical systems

Let's continue with our example of the body temperature measured with a thermocouple or a laser thermometer with a 0.1 °C precision. Our first measurement of the body temperature yielded (38.0±0.1) °C. Now let's repeat this measurement a number of times in various area of the body (which are left to your imagination). Let's say it then shows (38.1±0.1) °C, (38.0±0.1) °C, (38.3±0.1) °C, (37.9±0.1) °C, (38.2±0.1) °C, (38.1±0.1) °C, (38.1±0.1) °C, (39.8±0.1) °C. What is the actual body temperature then? Should we stick to a single measurement? Of course not. We have to make an histogram of the measured values, and study the distribution of the measurements (Fig. 3.1). We can then see that one of the values is clearly an outlier – something might have gone wrong there. What if we had done the measurement only once and only measured that value? We might have jumped to a very wrong conclusion, with possibly a very serious consequence like giving the wrong medicine.

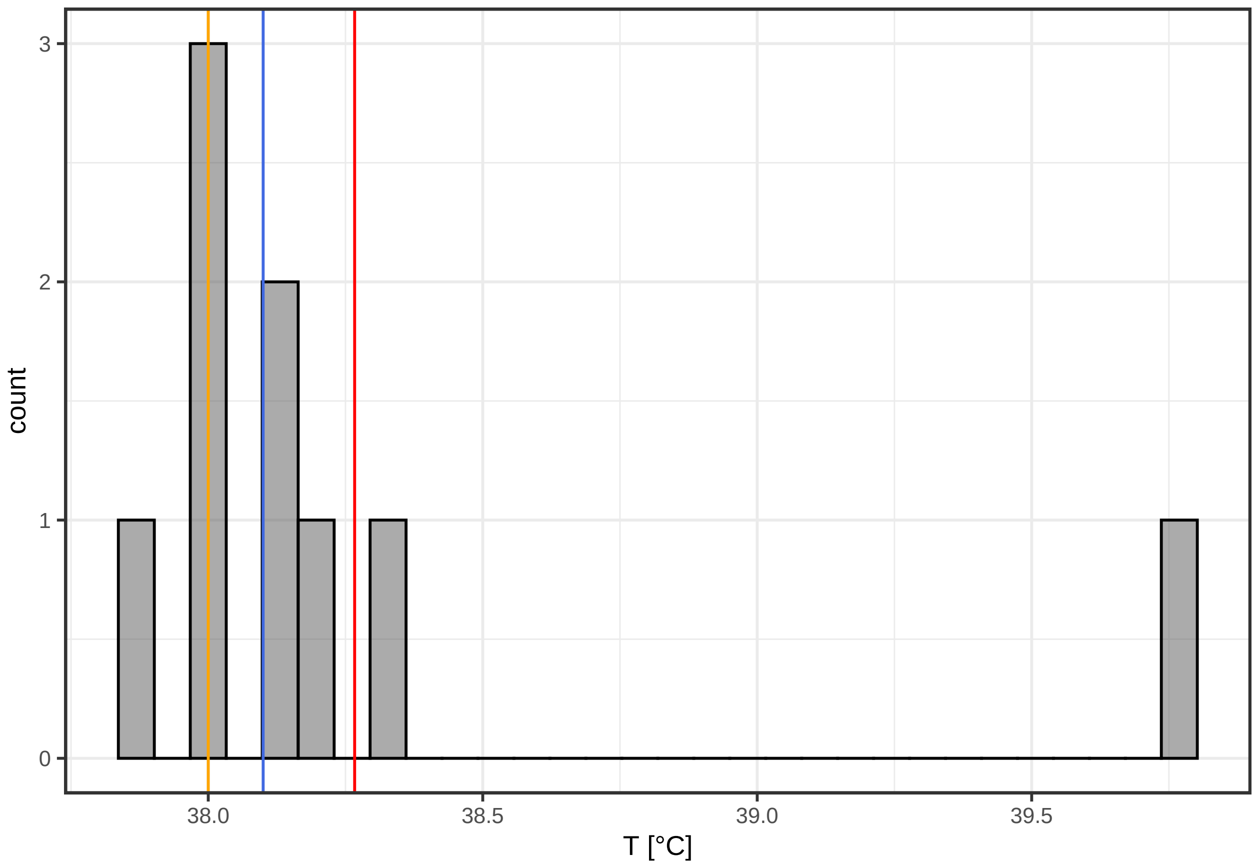

Figure 3.1: Histogram of the body temperature measurements. The red line is the mean value, the orange one is the mode and the blue one is the median.

With this example, we see that **a physical measurement is not absolute**. In fact, a physical measurement is an assessment of the **probability** that the physical value is within a certain range. In the case of our example, after removing the outlier for which we are certain that the measurement is wrong, it means that the measured body temperature has a high probability to be somewhere between 38.0 °C and 38.2 °C. In other (more general) terms, one could consider a measurement of a quantity X as a probability $P(x-\sigma < X < x+\sigma)$ that the quantity X has a value between $x-\sigma$ and $x+\sigma$. The **uncertainty** σ around the **mean value** x is usually given as the **standard deviation** of the distribution of measurements around the mean.

Since physical measurements are in fact **probabilities**, we **can** – and **must** – use **statistical tools** to characterize them.

References

[1] N. Ashcroft et al., *Solid State Physics*, Holt Saunders, New York, 1976.

[2] J. Doe et al., "An article title", Journal of Templates **90** (2022)