1

Data types

This chapter will ...

- Set out a system for describing different types of data.
- Explain why we need to identify the type of data with which we are dealing.

1.1 Does it really matter?

To open a statistics book with a discussion of the way in which data can be categorised into different types probably sounds horribly academic. However, the first step in selecting a data handling technique is generally identifying the type of data with which we are dealing. So, it may be dry, but it does have real consequences.

We will discuss three types of data. These go under a variety of names. The names that this book will use are (with common alternatives in brackets):

- Interval scale (Continuous measurement data)
- Ordinal scale (Ordered categorical data)
- Nominal scale (Categorical data)

Essential Statistics for the Pharmaceutical Sciences, Second Edition. Philip Rowe.

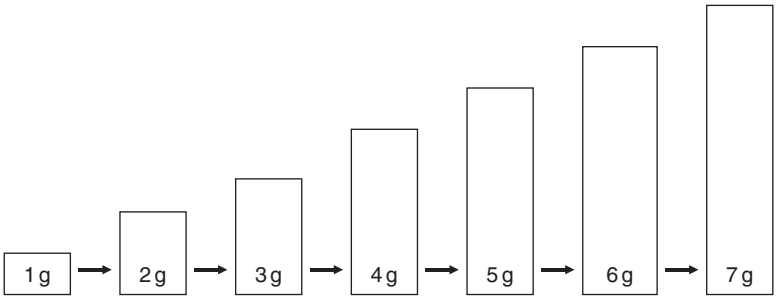

Figure 1.1 Interval scale data – a series of weights (1–7 g)

1.2 Interval scale data

The first two types of data that we will consider are both concerned with the measurement of some characteristic. 'Interval scale' (or what is sometimes called 'Continuous measured') data includes most of the information that would be generated in a laboratory. These include weights, lengths, timings, concentrations, pressures etc. Imagine we had a series of objects weighing 1, 2, 3 and so on up to 7 g, as in Figure 1.1.

Now think about the differences in weights as we step from one object to the next. These steps, each of one unit along the scale, have the following characteristics:

1. *The steps are of an exactly defined size*. If you told somebody that you had a series of objects like those described above, he or she would know exactly how large the weight differences are as we progressed along the series.

2. *All the steps are of exactly the same size*. The weight difference between the 1 and 2 g objects is the same as the step from 2 to 3 g or 6 to 7 and so on.

Because these measurements have constant sized steps (intervals), the measurement scale is described as a 'Constant interval scale' and the data as 'Interval scale'. Although the weights quoted in Figure 1.1 are exact integers, weights of 1.5 or 3.175 g are perfectly possible, so the measurement scale is said to be 'Continuous'.

1.3 Ordinal scale data

Again measurement is involved, but the characteristic being assessed is often more subjective in nature. It's all well and good to measure nice neat objective things like blood pressure or temperature, but it's also a good idea to get the patient's angle on

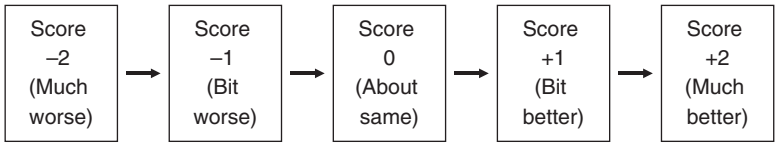

Figure 1.2 Ordinal scale data – scores for patient responses to treatment

how they feel about their treatment. The most obvious way to do this is as a score, of (say) –2 to +2 with the following equivalences:

–2 = Markedly worse
–1 = A bit worse
0 = About the same
+1 = A bit better
+2 = Markedly better

In this case (Figure 1.2) all we know is that if one patient reports a higher value than another, they are more satisfied with their outcome. However, we have no idea how much more satisfied he/she might be.

Since we have no idea how large the steps are between scores, we obviously could not claim that all steps are of equal size. In fact, it is not even necessarily the case that the difference between scores of –2 and 0 is greater than that between +1 and +2. So, neither of the special characteristics of a constant interval scale apply to this data.

The name 'Ordinal' reflects the fact that the various outcomes form an ordered sequence going from one extreme to its opposite. Such data is sometimes referred to as 'Ordered categorical'. In this case the data is usually discontinuous; individual cases being scored as –1 or +2 and so on, with no fractional values.

1.4 Nominal scale data

In this case there is no sense of measuring a characteristic; we use a system of classifications, with no natural ordering. For example, one of the factors that influences the effectiveness of treatment could be the specific manufacturer of a medical device. So, all patients would be classified as users of 'Smith', 'Jones', or 'Williams' equipment. There is no natural sequence to these; they are just three different makes.

With ordinal data we did at least know that a case scored as (say) +2 is going to be more similar to one scored +1 than to one scored 0 or –1. But, with nominal data, we have no reason to expect Smith or Jones equipment to have any special degree of similarity. Indeed the sequence in which one would list them may be entirely arbitrary.

Quite commonly there are just two categories in use. Obvious cases are Male/Female, Alive/Dead or Success/Failure. In these cases, the data is described as "Dichotomous".

Data types

Interval scale: Measurements with defined and constant intervals between successive values. Values are continuous.
Ordinal scale: Measurements using classifications with a natural sequence (lowest to highest) but with undefined intervals. Values are discontinuous.
Nominal scale: Classifications that form no natural sequence.

1.5 Structure of this book

The structure of this book is largely based upon the different data types. Chapters 3 to 16 all deal with the handling of continuous measurement data, with Chapters 17 to 20 focusing on categorical data; and then Chapter 21 covers ordinal data.

1.6 Chapter summary

When selecting statistical procedures, a vital first step is to identify the type of data that is being considered.

Data may be:

- Interval scale: Measurements on a scale with defined and constant intervals. Data is continuous.
- Ordinal scale: Measurements on a scale without defined intervals. Data is discontinuous.
- Nominal scale: Classifications that form no natural sequence.