Statistical Thinking for the 21st Century

Copyright 2020 Russell A. Poldrack

Chapter 2

Working with data

2.1 What are data?

The first important point about data is that data *are* - meaning that the word "data" is plural (though some people disagree with me on this). You might also wonder how to pronounce "data" – I say "day-tah" but I know many people who say "dah-tah" and I have been able to remain friends with them in spite of this. Now if I heard them say "the data is" then that would be a bigger issue...

2.2 Discrete versus continuous measurements

A *discrete* measurement is one that takes one of a set of particular values. These could be qualitative values (for example, different breeds of dogs) or numerical values (for example, how many friends one has on Facebook). Importantly, there is no middle ground between the measurements; it doesn't make sense to say that one has 33.7 friends.

A *continuous* measurement is one that is defined in terms of a real number. It could fall anywhere in a particular range of values, though usually our measurement tools will limit the precision with which we can measure; for example, a floor scale might measure weight to the nearest kg, even though weight could in theory be measured with much more precision.

It is common in statistics courses to go into more detail about different "scales" of measurement. The most important takeaway from this is that some kinds of statistics don't make sense on some kinds of data.

For example, imagine that we were to collect postal Zip Code data from a number of individuals. Those numbers are represented as integers, but they don't actually refer to a numeric scale; each zip code basically serves as a label for a different region. For this reason, it wouldn't make sense to talk about the average zip code, for example.