
Part 1: Terminology

In this topic there are a number of words you need to be familiar with the meaning of. In this section we will look at a few of them that will help with the understanding throughout the rest of the topic.

Populations and Samples

Population is a collection of all objects or individuals of interest that have properties that someone wishes to record. For example: "all people aged 18 and over who were living in New Zealand on 23 August 2018" or "all possible 15-watt LED light bulbs that could be produced by a manufacturing plant." Usually when we are looking at data we look at a **sample**. This is because it is not normally feasible to do a **census** (which looks at everything in a **population**). The sample is a group of objects, individuals or values that are chosen from the population. The idea is that this group will be representative of the population. Usually, the bigger the sample, the more reliable our estimates of the population (**point estimates**) are likely to be. When taking a **sample** it is important we avoid **bias**, which is anything that might cause favouritism to be shown towards one group or set of results. Sampling isn't the only time **bias** can occur, but is definitely one of them.

Variables

A **variable** is a measurement or a characteristic, for example weight or gender. **Explanatory** variables are variables that may provide information about another variable. **Response** variables are the variables that might be affected by the **explanatory** variable. **Control** variables are variables that are part of a controlled experiment that help ensure the results are valid.

Data Types

Qualitative (or categorical) data is data in which the values can be organised into distinct groups. These distinct groups (or categories) must be chosen so that they do not overlap and that every value belongs to one and only one group, and there should be no doubt as to which one. For example, eye colour.

Quantitative data is data in which the values result from counting or measuring. Measurement data is quantitative, as is whole-number data. There are two types of quantitative data, **discrete** and **continuous**. **Discrete** data is data where the data can only take on specific values, often whole numbers, for example, the number of people in a room. **Continuous** data is data which can be any value inside the range, for example the length of your foot.

Point Estimates

Because we are looking at a **sample** rather than the **population**, we need to make an **estimate** of what we think the **population** will be like. We can never be totally sure about this **estimate**, but it is our best guess based on our **sample**. An 'average' is a good example of a **point estimate** or **population parameter**, as based on our sample, we are **estimating** what we think the 'average' of the population will be. Usually, when referring to 'average', people are referring to the **mean**, which is calculated by adding all the values together and dividing by the number of points. 'Average' however can also refer to the **median**, which is the middle number when all the numbers are put in order, or the **mode**, which is the value that is most likely to occur.