Introduction to Statistics

Aaron McMurray

2022-06-09

Contents

Introduction to Statistics

Overview

This resource is intended to provide an introduction to the basics of statistics.

Contents

Chapter 1

Introduction

1.1 What is Statistics?

Statistics is all about the collection, organization, analysis, interpretation and presentation of data. Statistics is used everywhere from opinion polling in politics to predicting the prices of assets. There are two main branches of statistics: descriptive statistics and inferential statistics.

1.2 Descriptive Statistics

Descriptive statistics describes or summarises data that have been collected. Measures of central tendency such as (mean, median and the mode) and measures of dispersion (range, interquartile range and standard deviation) are the most important tools.

1.3 Inferential Statistics

Inferential statistical is used to make prediction about a population using information gathered about a sample. Inferential statistics involves hypothesis testing and regression analysis.

Chapter 2

Data Types and Levels of Measurement

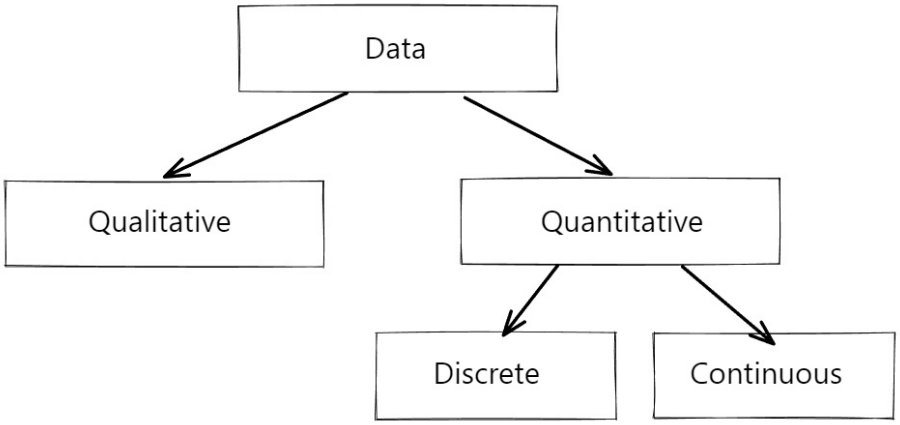
2.1 Types of Data

Data can be broadly categorised as **qualitative** (data relating to qualities or characteristics) or quantitative (numerical data relating to sizes or quantities of things).

We can further categorise **quantitative** data as being continuous or discrete.

Discrete data involves whole numbers that can't be divided because of what they represent (number of people in a class, number of cars owned). The number of people in a class cannot be 10.5 or 3.14. It must be a whole number because people are not divisible.

Continuous data can be divided and measured to some number of decimal places (height, weight, speed in miles per hour). A person's height can be any number (provided it lies within the range of possible human heights) and can be reported to any number of decimal places (150cm or 150.1cm or 150.12cm) depending on how accurate the measurement tool is.



There are also different **levels of measurement**.

2.2 Levels of Measurement

The levels of measurement describe how precisely variables are recorded. The different levels of measurement limit which statistics can be used to summarise data and which inferential statistics can be performed. These levels are:

- Nominal
- Ordinal
- Interval
- Ratio

2.2.1 Nominal

Nominal data is a type of data that is used to label variables. It can be categorised but not ranked (eye colour and gender for instance). The values grouped into these categories have no meaningful order. It is not possible to form a meaningful hierarchy of gender or eye colour.

The only measure of central tendency used with nominal data is the mode.

2.2.2 Ordinal

Ordinal data is another type of **qualitative data** that groups variables into descriptive categories. The categories used for ordinal data are ordered in some kind of hierarchical scale although the distance between those categories may be uneven or even unknown.

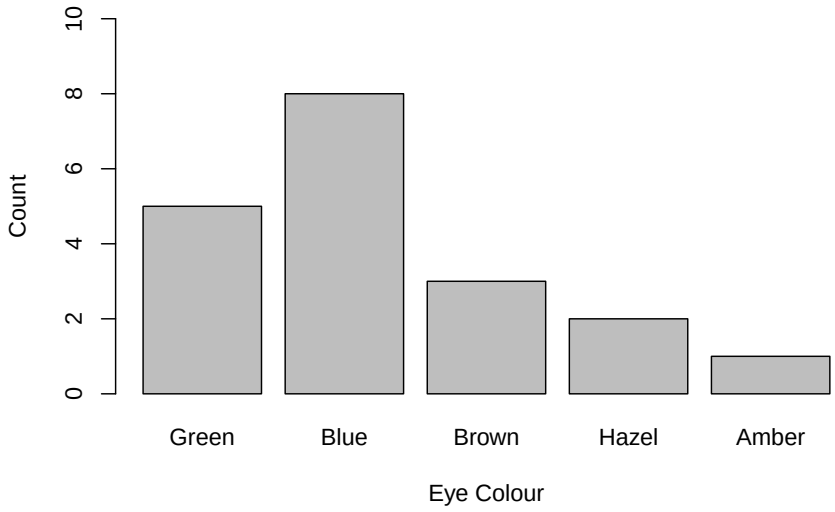


Figure 2.1: Eye colour is an example of nominal data.

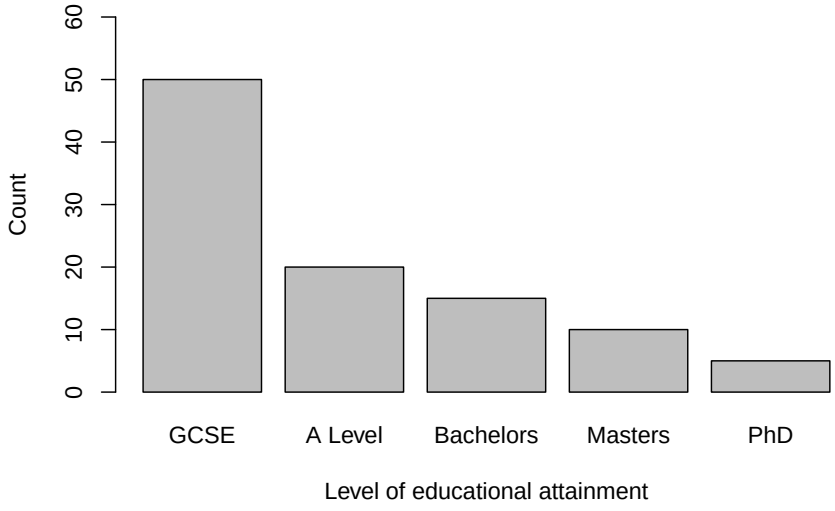


Figure 2.2: The highest level of educational attainment has a heirarchical scale but the distance between categories is unclear.

Ordinal variables often include ratings about opinions that can be categorised (strongly agree, agree, don't know, disagree, strongly disagree).

The descriptive statistics which can be used with ordinal data are the mode and the median.

Ordinal data can also be described with a measure of dispersion, namely, range.

2.2.3 Interval

Interval data is a type of quantitative data that groups variables into categories. Values can be ordered and separated using an equal measure of distance.

An example of interval level data is temperature data recorded in Celsius or Fahrenheit. The values on either scale are ordered and separated using an equal measure of distance (the distances between notches on a thermometer are always equally spaced).

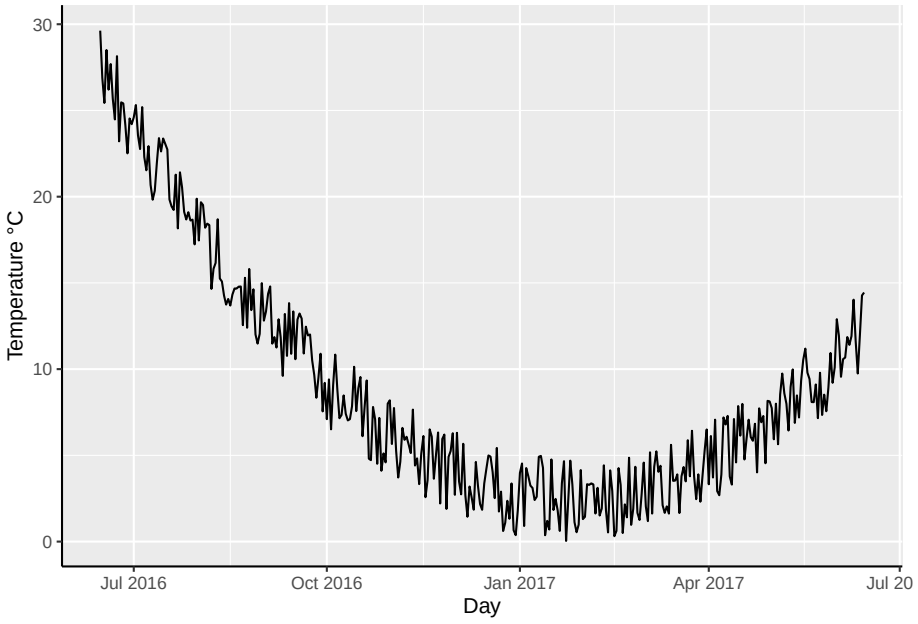


Figure 2.3: Temperature in Celsius is interval data. The values are ordered and separated by an equal interval. The distance between 0°C and 1°C is the same as the distance between 2°C and 3°C.

Mathematical operations can be carried out on this type of data, for instance, subtracting one value from another to find the difference. Interval data lacks a **true zero**.

True zero indicates a lack of whatever is being measured. The Celsius scale doesn't qualify as having a true zero since the zero point in a thermometer is arbitrary. When the Celsius scale was first created by Anders Celsius 0°C was selected to match the boiling point of water and a value of 100 °C was the freezing point of water. The scale was later reversed. Thermometers measure heat and at 0°C there is still heat, maybe not a great deal of it but heat is still measurable meaning 0°C is not a true zero. The thermodynamic Kelvin Scale has a true zero - where particles have no motion and can become no colder (there is a true absence of heat).

A range of descriptive statistics can be used to describe interval data. The measures of central tendency applicable to interval data are the **mode, median** and the **mean**. The measures of dispersion applicable to interval data are the **range, standard deviation** and the **variance**.

2.2.4 Ratio

Ratio data is a form of quantitative data. It measures variables on a continuous scale with an equal distance between adjacent values (weight, height). Ratio data has a true zero. Ratio data is the most complex of the four data types.

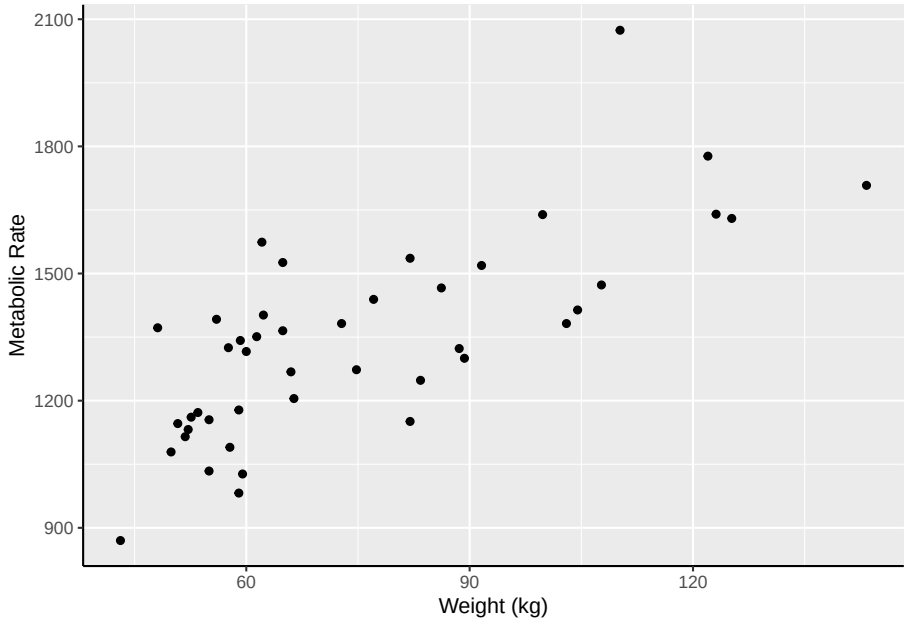


Figure 2.4: The scatterplot above shows metabolic rate plotted against body weight (kg). These are examples of ratio data.

Ratio data can be analysed with descriptive statistics including the **mode, median** and **mean**. **Range, standard deviation, variance** and the **coefficient of variation** can all be used to describe the dispersion of ratio data.

Chapter 3

Applications

Some *significant* applications are demonstrated in this chapter.

3.1 Example one

3.2 Example two

Chapter 4

Methods

We describe our methods in this chapter.

Chapter 5

Final Words

We have finished a nice book.