Segment: Descriptive Statistics

Topic: Types of Data

Table of Contents

Introduction

The first step in data analysis is learning to recognise the different types of data.

Different types of data may require different methods of organising and presenting the data. This topic introduces some common types of data and discusses the key differences between them.

Let us first look at an overview of this topic.

Types of data

Consider this.

- The identification number of the part that your company produces is 167402.
- Your annual income is S$167,402.

The number is the same in both cases.

But can you use the number the same way in both cases? No.

The income is a measured quantity to which you can add a number if that is an additional income, or from which you can subtract a number if that is an expense.

On the other hand, the identification number cannot be meaningfully used in any arithmetic calculation.

In this topic, you will study different types of data and the scale of measurement, and learn what can, and cannot, be done with each type of data.

Learning Objectives

At the end of this topic, you will be able to:

- identify the different types of data
- identify the different scales of measurement for measuring various types of data
- distinguish between types of quantitative data
- differentiate between data collected using various time scales.

1. Different Types of Data

Data can be broadly classified as:

- Quantitative
- Qualitative

Read below to find out more about these classifications.

Broad Data Classification

Quantitative

A quantitative variable is one that can be described by a number for which arithmetic operations, such as averaging, makes sense. Salaries for employees in a particular company might be an example of a quantitative variable. Since the variable is numerical or quantitative, calculations such as the average salary of employees is also meaningful.

Qualitative

A qualitative (or categorical) variable is a variable that records a quality. Variables such as gender are considered qualitative, as the possible values (male or female) are categories rather than numerical. If a number is used for distinguishing members of different categories of a qualitative variable, the number assignment is arbitrary.

It is mostly quantitative data that we will be working within later topics.

Exercise:

Below is an exercise to identify the difference between quantitative and qualitative variables.

A condominium is up for sale in the Boston area. Realtors who help sell the unit provide prospective buyers with the following information – asking price US$168,000, two bedrooms facing east, a washer, dryer and heater are included.

Question 1: Is the asking price of US$168,000 qualitative or quantitative information?

1. Quantitative

2. Qualitative

Question 2: Is the fact that the unit has two bedrooms qualitative or quantitative information?

1. Quantitative
2. Qualitative

Question 3: Is the fact that the unit faces east qualitative or quantitative information?

1. Quantitative
2. Qualitative

Question 4: Is the fact that the unit has a washer, dryer and heater qualitative or quantitative information?

1. Quantitative
2. Qualitative

2. Measurement Scales

Given some data, quantitative or qualitative, one should be clear about the scale in which it has been measured. There are four generally used scales of measurement. We shall note some important aspects of each scale listed here.

1. Nominal scale
2. Ordinal scale
3. Interval scale
4. Ratio scale

Read below to find out more about these four scales.

Scales of Measurements

1. Nominal Scale

In the **nominal scale** of measurement, numbers are used simply as labels for groups or classes. If our data set consists of blue, green and red items, we may designate blue = 1, green = 2,

and red = 3. In this case, the numbers 1, 2, and 3 stands only for the category to which a data point belongs. 'Nominal' stands for 'name' of category. Such data can only be used to **identify** or **classify** a person or thing.

2. Ordinal Scale

In the **ordinal scale** or measurement, data elements may be ordered according to their relative size or quality.

A good example of ordinal scale measurement is the **rank** of an athlete in a tournament. Such data can be used only for ordering from best to worst, or biggest to smallest, etc.

Consider a running race in which runners are ranked 1, 2, ... such that rank 1 is faster than rank 2 and so on. Looking at the ranks we can only say who is faster. Take the two runners whose ranks are 10 and 20... We can say that rank 10 is faster than rank 20.

But we cannot say that rank 10 was twice as fast as rank 20. Also, we cannot say that the difference between ranks 10 and 20 is twice as large as the difference between ranks 10 and 15.

3. Interval Scale

In the **interval scale** of measurement, the value of zero is assigned arbitrarily. Along major highways, miles or kilometres are marked prominently at various points.

If point A is marked 100 kilometres, and point B is marked 200 kilometres, does this mean point B is twice as far as point A is from where you are?

No. The reason is that you may not be at the 0 km point. When the zero marker is placed at some arbitrary point, we have an **interval** scale.

In this case, **only the intervals can be compared to produce meaningful ratios**. For example, if point C is marked 300 kilometres, then we can say that point C is twice as far from point A as point B is because the interval AC = (300 - 100) kilometre is twice as long as the interval AB = (200 - 100) kilometre.

Temperatures measured in Centigrade or Fahrenheit are another good example of an interval scale because zero degrees does not mean the absence of heat.

However, it is true that the heat required to raise the temperature of an object from 10 degrees Centigrade to 20 degrees Centigrade will be twice as much as the heat required to raise the temperature of the same object from 10 degrees Centigrade to 15 degrees Centigrade.

4. Ratio Scale

If two measurements are in ratio scale, then we can take ratios of those measurements. The zero in the scale is absolute zero.

Good examples of ratio scale measurements are weight, volume, income, cost, etc.

In these cases, **zero really means zero** and ratios can be taken straight away with the measured amounts. Weight of 20kg is indeed twice as heavy as 10kg.

3. Quantitative Data Variables

Among quantitative data, there are two important types of variables, namely,

1. Discrete variable
2. Continuous variable

Read below to find out more about these two types of variable.

Types of Quantitative Data Variables

1. Discrete Variables

Discrete variables are those that can assume only a countable or finite number of values. For example, the number of stocks in an investment portfolio, or the number of customers waiting in a queue are discrete variables as they can only assume the values 0, 1, 2, etc.

2. Continuous Variables

On the other hand, variables such as the return on a portfolio of stocks can vary continuously. For example, the annual return on a portfolio of stocks could be, say 8.765%.

In this case, our variable of interest could assume an infinite number of possible values, and thus it is classified as a continuous variable.

It is useful to differentiate between discrete and continuous variables, as they may need to be handled differently in many situations.

For example, to display the distribution of a discrete variable graphically, we might use a bar chart while for a continuous variable we might use a histogram.

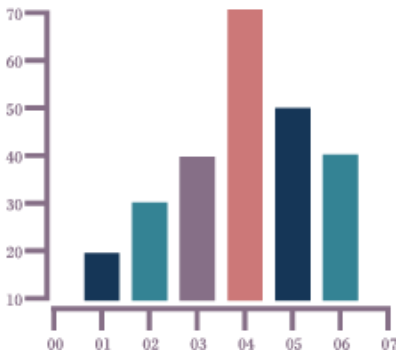

Fig. 1: Discrete Variable - Bar Chart

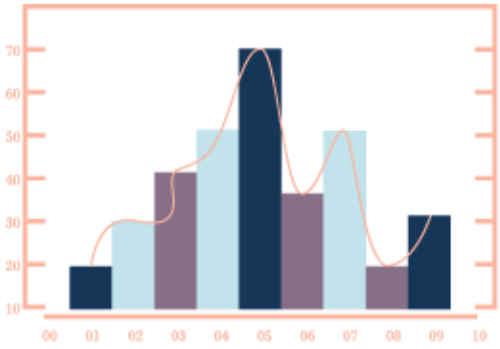

Fig. 2: Continuous Variable - Histogram

In the segment, “Probability and uncertainty”, we will explore the basic concepts involved in probability and in making decisions where uncertainty is involved. These concepts lay the groundwork for later work involving probability distributions, where we will need to treat discrete and continuous variables very differently.

4. Cross-Sectional and Time-Series Data

Data can also be classified in terms of whether the data was collected at a particular point in time (cross-sectional data) or whether collected at successive points in time (time-series data).

Read below to find out more about these two classifications.

Cross-sectional Data

An example might be data collected on price-earnings ratios for a group of companies quoted on the NASDAQ exchange, where the data was collected at the same point in time.

Time-series Data

An example of time-series data might be the closing value of a particular stock market index over the period of a year. The values collected represent time-series data, where the observations are collected over a period of time.

5. Summary

Here is a quick recap of what we have learnt so far:

- Data can be qualitative or quantitative.
- It is measured using four main types of scales:
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Among quantitative data, a distinction is drawn between discrete and continuous variables.
- Data can be also classified as cross-sectional or time-series, depending on whether it was collected at a particular point in time or over a period of time.