Kognity © 2023

4.4 Market research

Tool: Descriptive statistics

As part of the Business Management course, you will need to collect and analyse data. Descriptive statistics are tools that may help you present and interpret the data that you have collected. You should be familiar with most from your Mathematics course. Descriptive statistics studied in this course include:
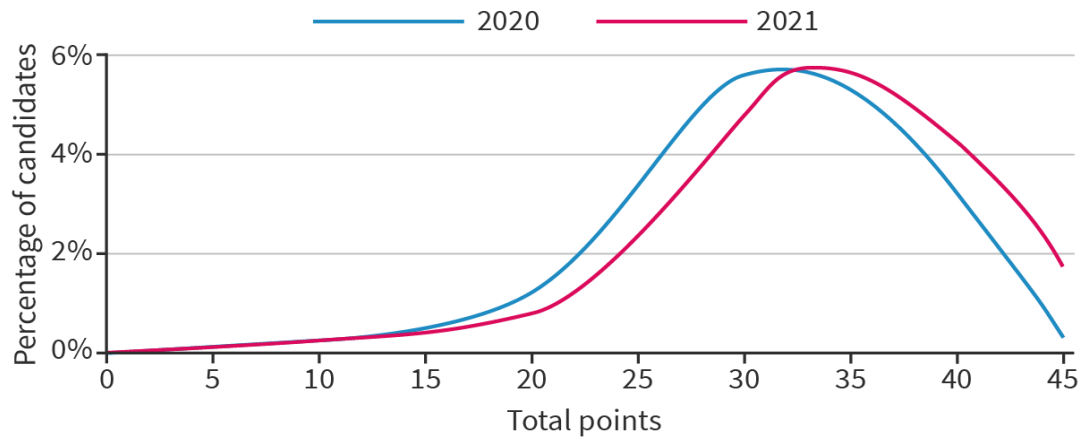
- mean
- mode
- median
- standard deviation
- pie charts
- bar charts
- infographics
- quartiles

Mean, mode and median

Mean is another word for average. The mean is found by adding together two or more numbers and dividing the total by the number of items. Mode refers to the most frequently occurring value from a set of values. Median is the middle value in a list of ordered numbers.

A mean or average is often used in reporting prices, such as the price of either a litre or a gallon of gas. Very high or very low numbers can skew the mean. Very often the median is then used to see the full picture. When talking about housing prices in an area, for example, it is more helpful to look at the median rather than the mean, as this ensures that one very expensive property does not increase the average.

Look at the distribution of IB diploma scores in **Figure 1**. Remember that the mode is always at the peak of a distribution of a given set of numbers, but this is not always true for the mean and median. For example, the mode of May 2021 IB diploma scores is 34 points, but the mean is 33.02 points.

Kog

Figure 1. Distribution of IB diploma scores.

Making connections

You should be familiar with mean, median, mode and distribution from your Mathematics course.
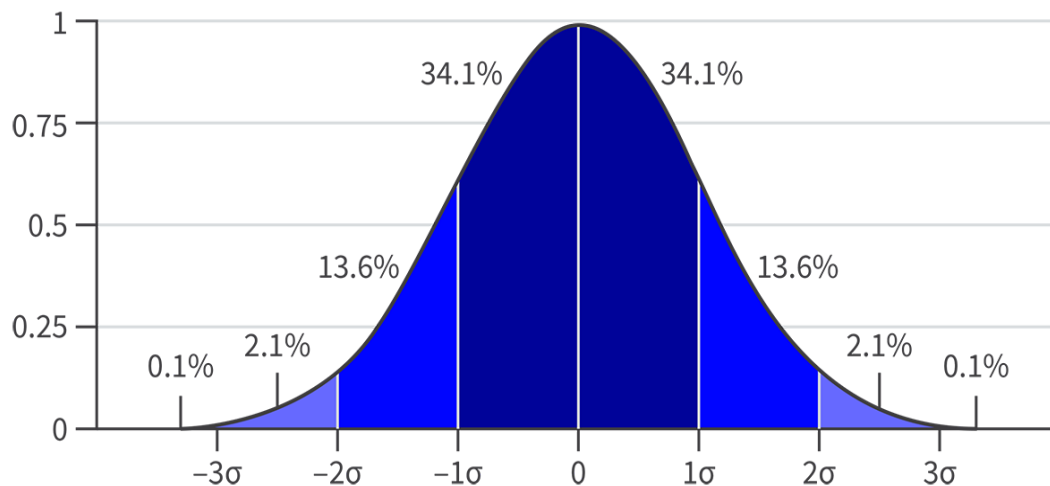
Standard deviation

Standard deviation – sometimes written using the Greek letter sigma (σ) – looks at the dispersion of data around its mean. A high standard deviation indicates that the data is spread out. A low standard deviation indicates that the data is clustered around the mean (see **Figure 2**).
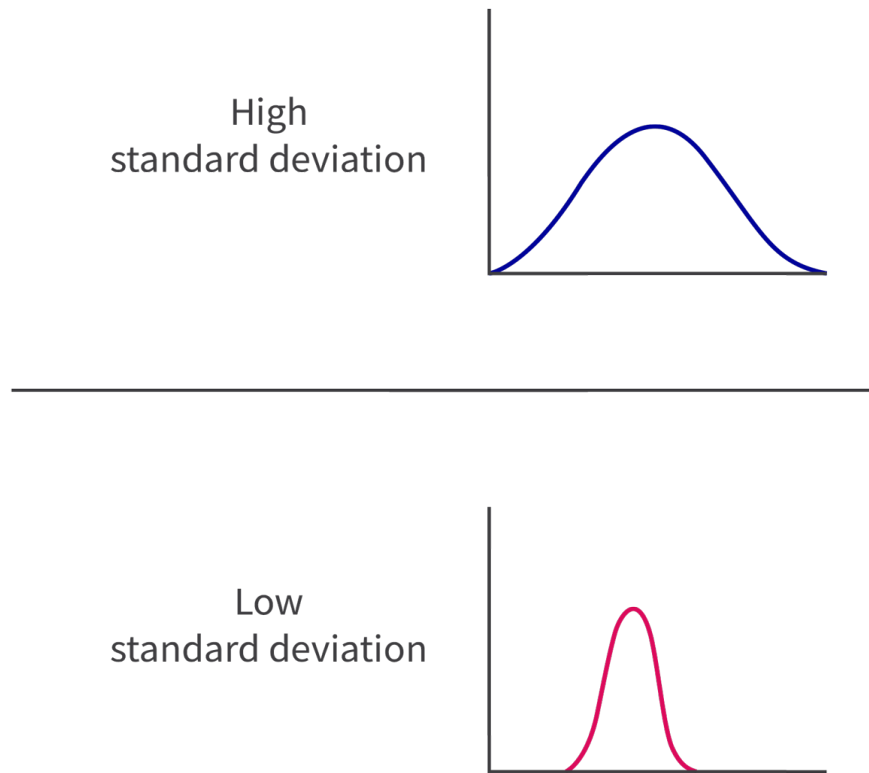
Figure 2. High and low standard deviation.

As shown in **Figure 3**, in a normally distributed set of data (a bell-shaped curve set):

- about 68% of data falls within one standard deviation of the mean
- about 95% of data falls within two standard deviations of the mean
- about 99.7% of data falls within three standard deviations of the mean

Figure 3. In a normally distributed set of data, 68.2% of data falls within one standard deviation of the mean.

One of the key standard deviation applications in business is to analyse the price changes of a share of a publicly held company and to understand the statistical likelihood of an event taking place.

You will not need to calculate the standard deviation in your Business Management exam, but if you are interested in how it is calculated, the following video will take you through it.

Video 1. A step-by-step explanation of how standard deviation is calculated.

Activity

Learner profile: Inquirer
Approaches to learning: Thinking skills (transfer)

Company XYZ's annual revenue data is shown in **Table 1**.

Table 1. Company XYZ's annual revenue.

Year	Revenue
2020	$181 billion
2019	$160 billion
2018	$136 billion
2017	$110 billion
2016	$89 billion
2015	$74 billion
2014	$65 billion

1. What would the mean revenue for the years 2014 to 2020 be?

2. What would the median revenue for the years 2014 to 2020 be?

3. The standard deviation for the set of data above is 40.87. Based on this standard deviation, if the 2021 revenue were to be $450 billion, would this be characterised as an outlier?

> ## International Mindedness
>
> Be mindful of regional differences when presenting and interpreting data. For example:
>
> - In some countries, a comma is used to separate thousands; in other countries a full stop or a space is used. In the USA, for example, the number ten thousand would be written as 10,000, while in Germany it would be written as 10.000. In Russia the correct way of writing the number would be 10 000. This Kognity book uses a space to separate thousands.
> - European countries mostly use the metric system (metres, kilometres, kilograms, etc.), while the USA uses the imperial system (feet, miles, pounds, etc.).

Bar Charts and pie charts

A bar chart is a chart with rectangular bars showing the values represented (see **Figure 4**). Bar charts are generally used to compare different occurrences of the same event or item. For example, you can use a bar chart to compare price changes of one product over time as well as to compare prices of similar products.

A pie chart is a circular graph in which segments of the circle represent percentages of the total (see **Figure 4**). Pie charts are generally used to represent percentages of a whole, such as the breakdown of sales by product.

IB Business Management HL scores – May 2021
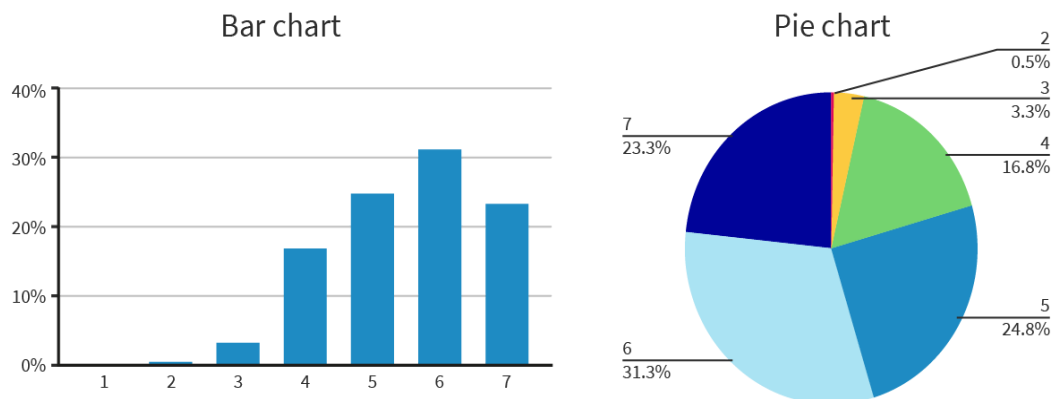
Figure 4. A bar chart is often used to compare, while a pie chart is used to show parts of a whole.

Activity

Learner profile: Inquirer
Approaches to learning: Research skills (information literacy)

You may or may not have used Microsoft Excel or Google Sheets (or a different spreadsheet program) in your studies to create bar charts and pie charts. The skill of creating your own graphs using technology will be of great use for presenting data in your internal assessment.

1. Company XYZ's annual revenue data is shown again in **Table 2**.

Table 2. Company XYZ's annual revenue.

Year	Revenue
2020	$181 billion
2019	$160 billion
2018	$136 billion
2017	$110 billion
2016	$89 billion
2015	$74 billion
2014	$65 billion

Use the data in **Table 2** to create a bar chart. Remember to label the axes correctly and fully (for example with the units and the currency), to use the appropriate units and to give the bar chart a title. Ensure that the title is specific (for example, include the name of the company, the year range and the units).

2. Company XYZ's revenue composition in 2020 is as follows:

Advertising revenue: 80%
Cloud revenue: 8%
Other revenue: 12%

Use the data above to create a pie chart. Remember to give the pie chart a title and to label each segment.

Infographics

An infographic is a graphic representation of information (see **Figure 5**). It is used to show research findings or to break down data in a visual way. You should be very careful when relying on infographics to convey information. Infographics may be artistically represented but may often not be drawn to scale.
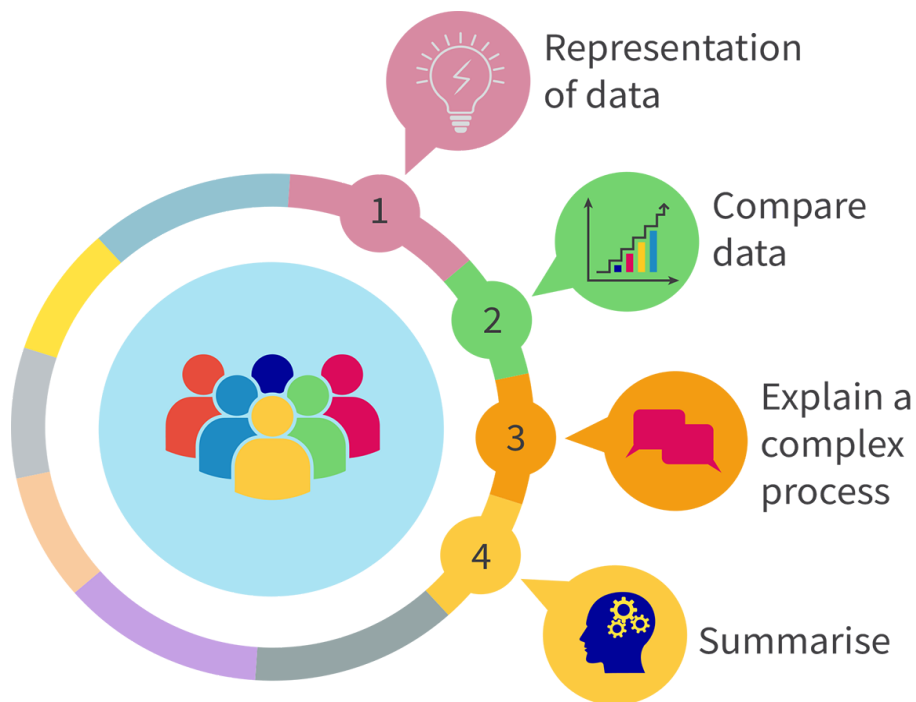

Figure 5. Infographics can be a useful way to visualise data, but the data may not necessarily be drawn to scale.

Quartiles

Quartiles result from dividing a set of numbers into quarters (see **Figure 6**). For example, a country's income data might be represented by looking at:

- the lowest 25% of income earners
- the second lowest 25% of income earners
- the second highest 25% of income earners
- the highest 25% of income earners
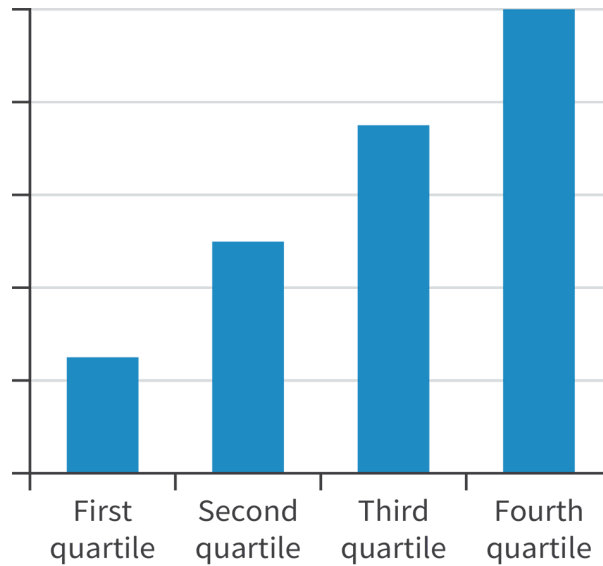
Figure 6. Quartiles result from dividing a set of numbers into quarters.

Problems with data presentation and interpretation

'*There are three kinds of lies: lies, damned lies, and statistics.*'
(Attributed to former UK Prime Minister, Benjamin Disraeli.)

Descriptive statistics are an excellent way to understand the meaning of data. However, interpreting graphic representations of data takes some skill. When looking at data, you need to pay close attention to the following:

- **The title of the diagram.** Take the time to read and understand the title of the diagram. Titles may be misleading, sometimes intentionally.

- **Axes and scale.** The two diagrams in **Figure 7** look very different. At first glance the first graphic appears to show significant growth and the second graphic appears to show no growth. In fact, the two diagrams show the same data but with different scales chosen for the *y*-axis.
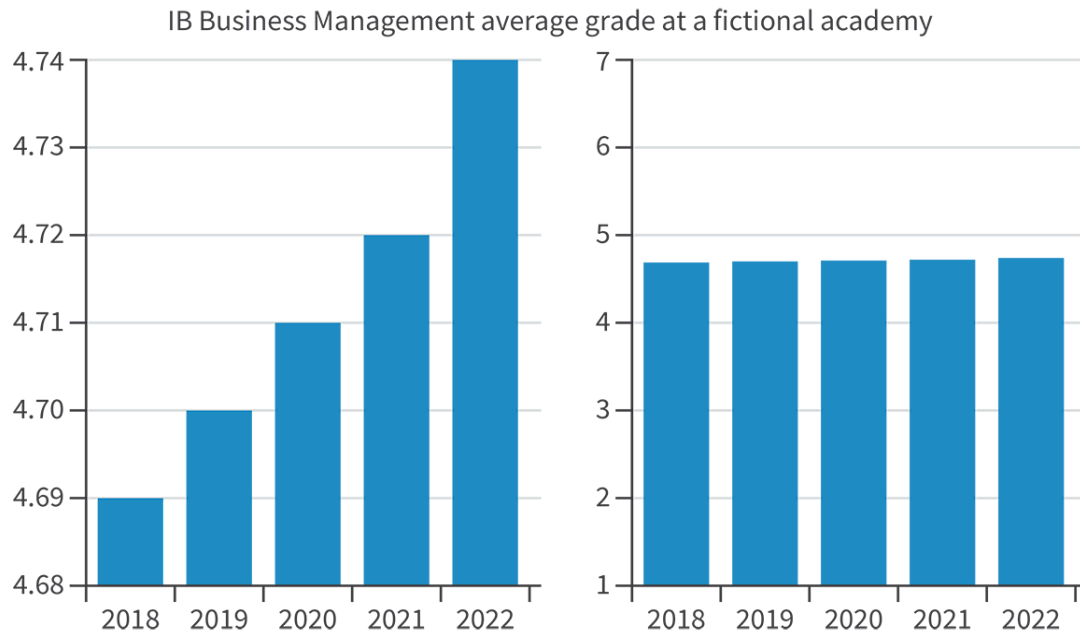
Figure 7. Using the same data but changing the scale of the vertical axis.

- Year range of data. Sometimes data is 'cherry picked' by year. This means that only the data that is convenient to the writer is presented. You should always pay close attention to the *x*-axis chosen for data over time (see **Figure 8**).
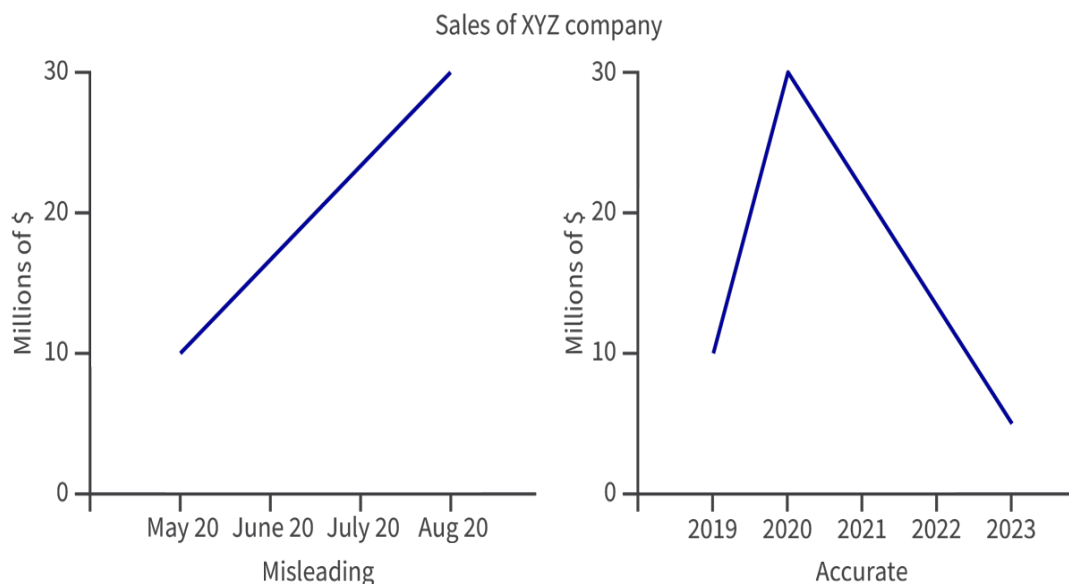
Figure 8. Using the same data but changing the scale of the horizontal axis.

- **Correlation versus causation.** Correlation is a statistical term outlining a relationship between two variables, which may or may not be a causal one. Remember that just because events are correlated, it does not mean that one is caused by the other (see **Figure 9**).
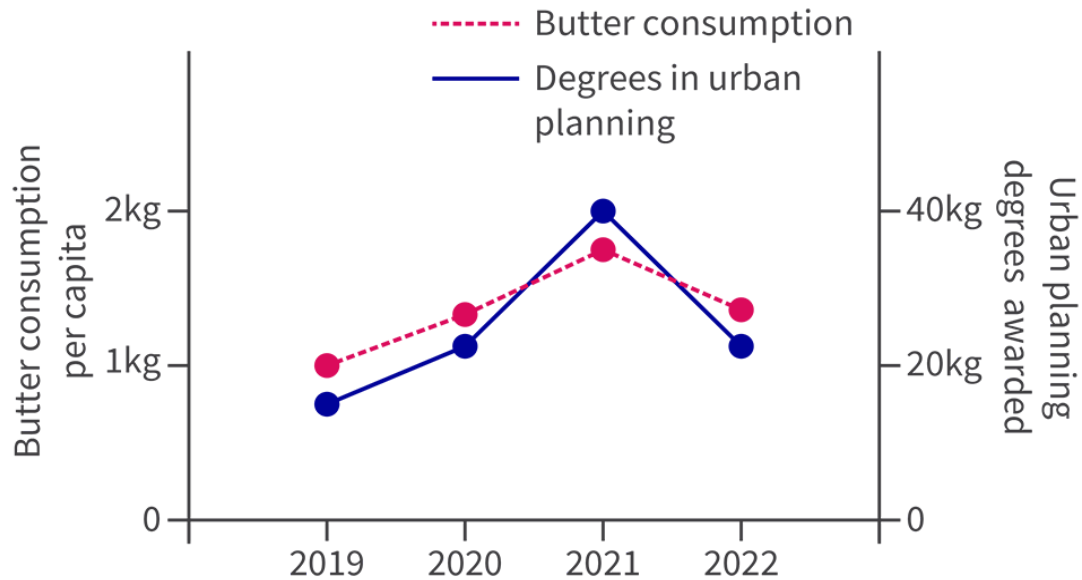
Figure 9. Correlation, not causation.

Watch the following video to understand how statistics may be used in a misleading way.

How statistics can be misleading - Mark Liddell

Video 2. Examples of how statistics can be misleading.

Concept

Ethics

As seen in the examples above, it is quite simple to present data in misleading ways, and you will need to be careful when relying on data to make conclusions. For this reason, presentation of data carries with it ethical considerations.

Exam tip

As part of this course, you will need to write an internal assessment, which is a research project about a real business issue or problem facing an organisation.

You will need to select your own primary and/or secondary sources. If primary research is conducted, you will also need to present data in a way that allows you to analyse the business and answer the research question.

Kog