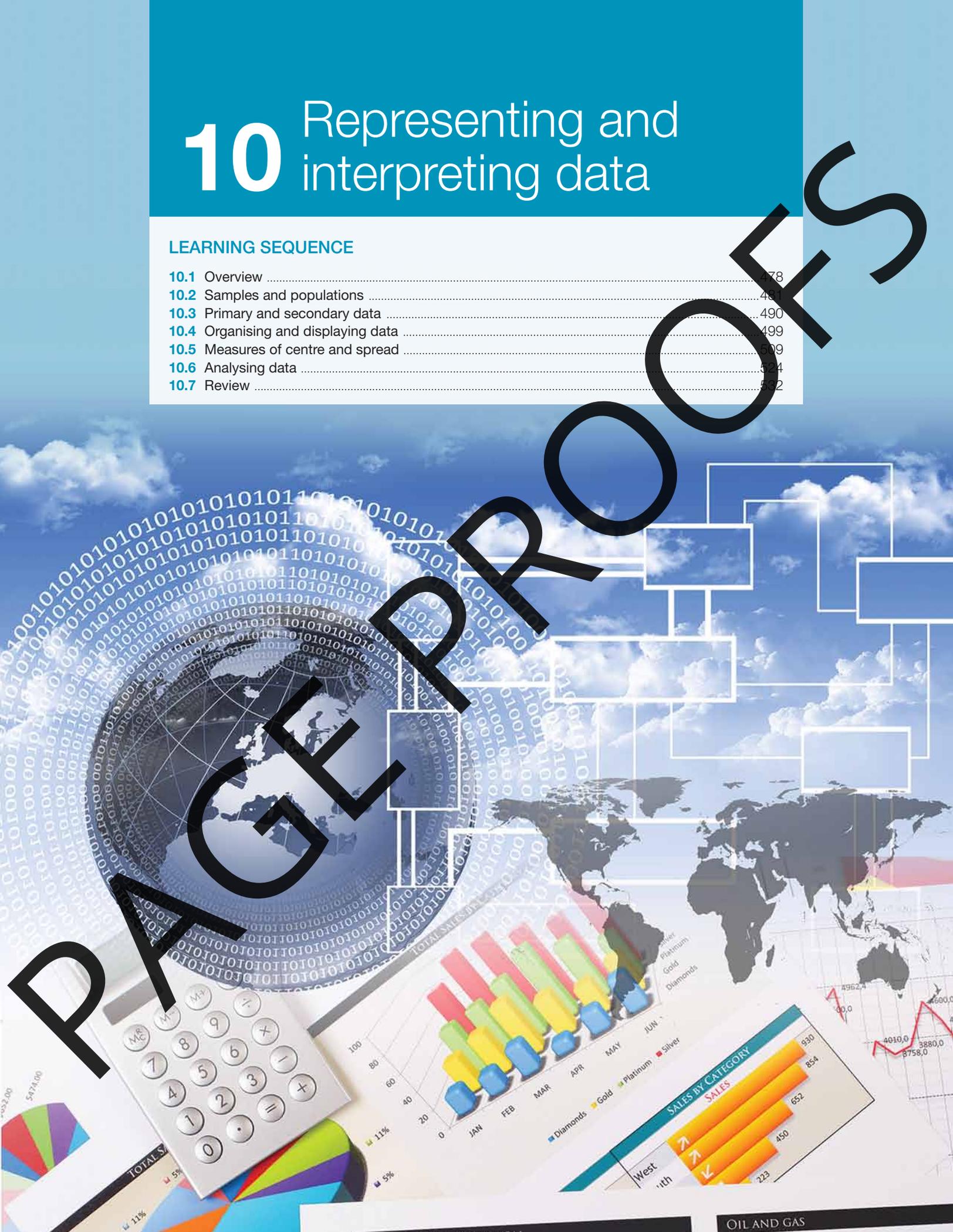


10 Representing and interpreting data

LEARNING SEQUENCE

10.1 Overview	478
10.2 Samples and populations	481
10.3 Primary and secondary data	490
10.4 Organising and displaying data	499
10.5 Measures of centre and spread	509
10.6 Analysing data	524
10.7 Review	532



10.1 Overview

Why learn this?

We see statistics and data everywhere. We read them in newspapers, hear them on the TV and see them quoted on social media. It is very important that we can understand and interpret the data that we see. If you read an article that uses the mean or median or shows a graph to describe data, you need to be able to make sense of the information. Imagine you hear on the TV that the mean age of a social media user is 40 years. If you have no idea what the mean is and what it tells us, then this information will not tell you anything. This topic will help you to understand and make sense of statistical data and graphs.

Statistics are used by most professions. Sporting organisations use statistics to analyse game data, advertising agencies to try to make us buy certain products, the media in reporting, economists to analyse share markets, and organisations to describe their performance. In any occupation, you would need to read and interpret statistics at some point, so it is important you can do this easily.

Where to get help

Go to your learnON title at www.jacplus.com.au to access the following digital resources. The Online Resources Summary at the end of this topic provides a full list of what's available to help you learn the concepts covered in this topic.

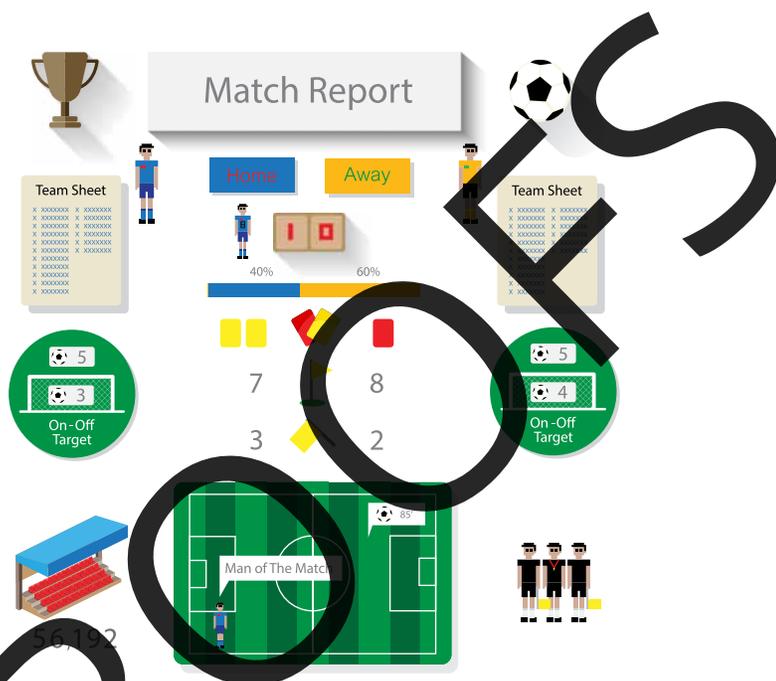
Video
eLessons

Digital
documents

Interactivities

eWorkbook

Fully worked
solutions
to every
question



Complete this pre-test in your learnON title at www.jacplus.com.au and receive **automatic marks**, **immediate corrective feedback** and **fully worked solutions**.

- Andy asked each person in his class what their favourite sport was. State the type of data that Andy collected.

- Consider the following set of data.

3, 3, 3, 3, 4, 5, 9, 11, 13

Calculate the:

- mode
- range
- median.

- MC** From the following list, select all the questions that should not be used in a questionnaire about school uniforms.

- The present uniform is pretty ugly. Don't you agree we should change it?
- To which age group do you belong?
- Do you like the skirt, tie and blazer?
- How do you feel about the price of the uniform?
- Are you in favour of changing the school uniform?

- MC** A sample of people were asked about the number of hours per week they spend exercising. The data was put into a grouped frequency table with class intervals of:

0 — < 4
4 — < 8
8 — < 12 and so on.

Which kind of graph would be the best choice for presenting this data?

- Bar chart
 - Histogram
 - Stem-and-leaf plot
 - Dot plot
 - Back-to-back stem-and-leaf plot
- A group of friends recorded how many hours they each spent watching television in a week. Their results are shown in the following table.

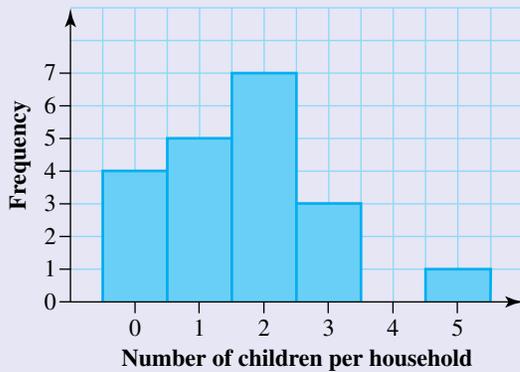
Value	Frequency
0	5
1	7
2	1
3	1

- Calculate the mean for the data, correct to 2 decimal places.
- Identify the median for this data.



6. **MC** This graph shows the number of children per household in a survey of 20 households.

Children per household in 20 households



Select the statement that is false.

- A. The modal number of children in a household is 2.
- B. There are 20 children in total.
- C. The range of the number of children is 5.
- D. There are 33 children in total.
- E. No household in the survey has 4 children.

7. Calculate the mean, \bar{x} , for this set of data:

3, 3, 3, 3, 4, 5, 7

8. Consider the following data:

0, 8, 9, 10, 10, 12, 13, 13, 14

Determine which measure of centre would be most appropriate in this situation.

9. **MC** The mean, mode, median and range are calculated for the following data:

2, 7, 8, 9, 10, 10, 12, 13, 14

The value of 2 is considered an outlier (or an anomaly) and is rejected. The measures are recalculated. From the following, select the true statement after the recalculation.

- A. The mode changes.
- B. The mean changes.
- C. The range remains unchanged.
- D. The median remains unchanged.
- E. None of the above is true.

10. Consider the following frequency distribution table.

Value	Frequency
0	8
1	7
2	4
3	2

What is the mean of this data?

11. A group of 22 Year 8 students measured their heights and recorded them in a grouped frequency table.

Height (cm)	Frequency
150– < 155	2
155– < 160	3
160– < 165	4
165– < 170	7
170– < 175	5
175– < 180	1



Calculate the mean height for the group of Year 8 students, correct to 1 decimal place.

12. **MC** From the following list, choose which methods of data collection can be used for primary data. Select all that apply.
- A. Survey
 - B. Questionnaire
 - C. Observation
 - D. Reading a report
 - E. Researching on the internet
13. The following values represent the number of hours worked in a week by 10 workers in a store:
34, 44, 28, 38, 36, 39, 10, 24, 28, 36
- If all of these values were doubled, determine the effect this would have on the mean.
14. Determine the sequence of six numbers with a median of 5, a mode of 5, a range of 2 and a mean of 4.5. Write the numbers from lowest to highest.
15. A data set containing 7 pieces of data has the highest value of q and the lowest of p . The mean is m . If another value $(m + 2)$ is added to the data set, determine by how much the value of the mean would increase.

10.2 Samples and populations

LEARNING INTENTIONS

At the end of this subtopic you should be able to:

- understand the difference between a sample and a population
- understand the various methods of collecting data, such as a survey, census or questionnaire
- select an appropriate sample and know if it is biased or not.

10.2.1 Collecting data

- **Data** is information that has been collected for the purpose of analysis.
- Data collected from the whole group being studied is known as a **population**.
- Data collected from a selection of a larger group is known as a **sample**.

- Different organisations collect different types of data. For example, at a cricket match, some of the **statistics** gathered for a batsman are: time spent batting, the number of balls faced, the runs off a particular delivery, where the ball was hit, the number of 4s or 6s hit, and so on.
- Once collected, the data can be organised, analysed and interpreted.
- A **survey** is the process of collecting data. If every member of a target population is surveyed, the process is called a **census**.
- A census is conducted in Australia every 5 years to obtain an accurate profile of Australians. On census night, each household in Australia is required to complete a detailed online questionnaire or booklet containing a series of questions relating to age, marital status, employment, income, housing, education, modes of transport and so on. This allows the government to analyse the population and make decisions on how to improve services.
- Due to limitations in time, cost and practicality, instead of a census, in many cases a sample of the population is selected at **random** (not in any particular order or pattern) to prevent **biased** (leaning in a favoured direction) results.



Simple random sampling



DISCUSSION

Discuss the practicalities of collecting data through a census compared to collecting data of a sample of the population. Consider the size of the population when taking a census, particularly in countries with large populations such as India or China.

WORKED EXAMPLE 1 Identify whether information was obtained by census or survey

In each of the following, state whether the information was obtained by census or survey.

- A school uses a roll to count the number of students who are absent each day.
- Television ratings are determined by 2000 families completing a questionnaire on the programs they watch over a one-week period.
- A battery manufacturer tests every hundredth battery off the production line.
- A teacher records exam marks for each student in her class.

THINK

- Every student is counted at roll call each morning.
- Not every family is asked to complete a ratings questionnaire.
- Not every battery is tested.
- The marks of every student are recorded.

WRITE

- Census
- Survey
- Survey
- Census

- Data can be collected using a variety of techniques. Three common methods are:
 - observation
 - survey
 - experiment.

eles-4435

10.2.2 Surveys

- Collecting data by survey is the form of data collection most frequently used.
- A survey is administered with the aid of a questionnaire.
- Methods used to conduct surveys include:
 - personal interviews, where the interviewer asks prepared questions
 - telephone interviews
 - self-administered questionnaires, which are usually completed online or posted to individuals.

Questionnaires

- When designing a **questionnaire**, keep in mind that the collected data should be in a form that is easy to analyse. Questions can be **open** or **closed**.
- Open questions are those for which the respondent has no guided boundaries within which to answer. Questions that belong to this class include:

‘Who is your favourite singer?’
 ‘What is your favourite food?’

The problem with open questions is that their answers are often difficult to classify and analyse.

- Closed questions are those that the respondent must answer within a category.

The question about food above could be rewritten as:

‘Which one of the following foods do you prefer most?’

- Meat
- Seafood
- Poultry
- Vegetables
- Fruit
- Answers to closed questions are easier to analyse than answers to open questions. It must be noted that options such as:
 - None of the above
 - Don’t know
 should be avoided, if possible, as they provide the respondent with a ‘way out’.



WORKED EXAMPLE 2 Analysing an open question

Ten students were asked the open question:

‘What do you dislike about your school uniform?’

These were their responses:

1. It's too old-fashioned.
2. It costs too much.
3. It's just yucky.
4. It's a bad colour.
5. The girls aren't allowed to wear pants.
6. The colour is drab.
7. The skirt is too long.
8. It would be better to have it all one colour rather than three different colours.
9. The shoes are too uncomfortable.
10. We should be allowed to wear leggings in winter.

Classify the responses into categories to identify the two main reasons given.

THINK

1. Read the responses and identify 2 or 3 common themes.
2. Classify the responses.
3. Identify the main reasons.

WRITE

Two main categories are apparent.

1. Style
2. Colour

A third category is not apparent.

Style

Responses 1, 5, 7, 9, 10

Colour

Responses 4, 6, 8

Responses 2 and 3 can be in a miscellaneous category.

The main reasons students dislike the uniform are its style and colour.

- It can be seen from Worked example 2 that it is probably worth taking the time to write closed questions for a questionnaire, as this reduces the number of answers that are difficult to classify.

10.2.3 Planning a questionnaire

eles-4437

- To prepare good questions, keep the following points in mind.
 1. The questions should flow smoothly from one to the next.
 2. Introductory remarks should be included, outlining the aim and purpose of the questionnaire, along with any necessary instructions.
 3. Jargon, slang and abbreviations should be avoided.
 4. Do not ask questions that are vague or ambiguous.
 5. Avoid bias and emotional language.
 6. Avoid double-barrelled questions (i.e. questions that have two purposes or meanings).
 7. Do not ask leading questions (those that lead to an expected response).
 8. Make sure your questions are capable of being answered by your respondents.
 9. Avoid questions with double negatives.
 10. At the conclusion, thank the respondent for answering.



WORKED EXAMPLE 3 Writing good questions

Identify the issues with the following questions, then rewrite each so that the meaning is clear, unbiased and understandable.

- Don't you agree that the school uniform should be changed?
- Didn't you know that students are not allowed to go to the school canteen at morning break?
- Those poor, distraught animals in the zoo should be running wild and not restricted in cages. Don't you agree?
- How old are you?



THINK

- This is a leading question — leading the respondent to agree. Avoid language like 'don't you think that' when writing questions.
- This question has two negative parts ('didn't' and 'not'), which makes it confusing.
- This question plays on the emotions of the respondent by using language such as 'poor' and 'distraught'.
- People often don't like disclosing personal details.

WRITE

- Are you in favour of changing the school uniform?
- Are you aware of the rule banning students from the school canteen at morning tea?
- What's your opinion about keeping animals in cages at zoos?
- To which age group do you belong?
 - Less than 10 years of age
 - 10–20 years of age
 - 21–50 years of age
 - Over 50 years of age

10.2.4 Selecting samples

eles-4438

- When a sample is selected from a population, care must be taken to ensure that the composition of the sample is as close as possible to that of the population.
- One of the simplest ways to select a reliable sample is to use a random sampling technique.
- A way in which this can be achieved is to number each individual item (element) of the population, then randomly select items for the sample by using a random number generator.



Digital technology

Random number generators can be found online and are standard features in Excel, graphing calculators and most scientific calculators.

To generate a random number between 0 and 1:

- using a calculator, look for a command called **rand**, **Ran** or similar. For example, to generate a random number between 0 and 30, type $30 \times \text{rand}$ followed by ENTER.
- using Excel, type =RAND() into an empty cell and press ENTER.

To generate a random integer:

- using a calculator, look for a command called **randint**, **RanInt** or similar. These commands are usually followed by the lower and upper limits of your random integer. For example, to generate a random integer between 0 and 30, type $\text{randint}(0, 30)$ and press ENTER.
- in Excel, the command =RANDBETWEEN can be used to generate random integers.

Note: Command names vary slightly between calculator brands.



WORKED EXAMPLE 4 Generating random numbers

As part of their quality control program, staff members at Perplexing Puzzles randomly select five puzzles for testing from a batch of 50 coming off a production line. Use a random number generator to select the five puzzles to be tested.

THINK

1. Assign a number from 1 to 50 to each puzzle on the production line, according to its order.
2. Use a random number generator to display five random numbers.
In a spreadsheet, type =RANDBETWEEN(1, 50) into a cell and press ENTER. Repeat this process until you have generated five unique numbers.
Note: Reject any repeated values.
3. List the puzzles whose numbers correspond to those selected in step 2 and interpret the results.

WRITE

- 1 — First puzzle on the production line
 - 2 — Second puzzle on the production line
 - 3 — Third puzzle on the production line
 - ...
 - 50 — Fiftieth puzzle on the production line
- Random numbers are: 48, 2, 8, 42 and 21
Note: Random numbers will vary between different devices.

The puzzles numbered 48, 2, 8, 42 and 21 will be selected for testing.

10.2.5 Biased samples

- If a sample is not representative of the population, it is said to be biased.
- To avoid bias, the size of the sample should be approximately $\sqrt{\text{population size}}$.

WORKED EXAMPLE 5 Determining a biased sample

Comment on the selection of the sample in the following cases.

- Members of the men's basketball team were interviewed regarding the height of the mirrors in the gym changing room.
- Spectators at a music concert were interviewed regarding the sporting facilities in the town.
- Of the 100 people in the bushwalking group, 5 were interviewed about increasing the club's fees.

THINK

- Are members of the men's basketball team of typical height compared to those who use the gym?
- Spectators at a music concert might not use sporting facilities very frequently.
- This sample is too small.

WRITE

- Members of the men's basketball team would be taller than most of the people who use the gym changing room. They would prefer the mirrors to be higher. This is a biased sample.
- This is a biased sample as it does not include those who would make the greatest use of the sporting facilities.
- This sample of 5 from the group of 100 is too small to represent the views of the whole group.

COLLABORATIVE TASK: Designing a survey

- Working in pairs, think of a topic on which you would like to gather your classmates' opinions. Some suggestions are: pets, movies, interests, favourite phone app, subjects studied.
- Think of five questions you would like to ask your classmates about this topic, ensuring that you collect both categorical and numerical data.
- When writing your questions, you should consider what sorts of answers you want. Do you want questions requiring *yes* and *no* answers, or would you like questions that require people to make a judgement on a scale? Questions should be easy to understand and must be relevant to your topic.
- Write down how you intend to collect the data, the type of data you expect to collect and how you think it is best to represent this data. Think about the possible responses you may get.
- As a class, listen to and offer constructive advice on the questions that each pair has developed. Feedback may consist of comments about the type of responses you might expect, how you could collect the data, whether the questions make sense and whether they are relevant to the topic.
- Once you have received your feedback, you may wish to reword some of your questions and rethink the way you collect the data.
- Collect and store the data. You will need to use it again during this topic.



-  **eWorkbook** Topic 10 Workbook (worksheets, code puzzle and project) (ewbk-1941)
-  **Interactivities** Individual pathway interactivity: Samples and populations (int-4451)
 - Collecting data (int-3807)
 - Questionnaires (int-3809)
 - Planning a questionnaire (int-3810)
 - Selecting samples (int-3811)
 - Biased samples (int-3812)

Exercise 10.2 Samples and populations

Individual pathways

PRACTISE
1, 2, 3, 5, 6, 10, 13

CONSOLIDATE
4, 7, 9, 11, 14

MASTER
8, 12, 15

To answer questions online and to receive **immediate corrective feedback** and **fully worked solutions** for all questions, go to your learnON title at www.jacplus.com.au.

Fluency

1. Copy and complete the following sentence.
When we obtain data from the whole population, we conduct a _____; however, a survey obtains data from a _____ of the population.
2. A school conducts an election for the new school captain. Every teacher and student in the school casts a vote. Is this an example of a census or a survey? Explain your answer.
3. A questionnaire is conducted by the school council to see what library facilities the community needs, and 500 people who live in the community participate. State whether this is an example of a census or a survey.
4. **WE1** For each of the following, state whether a census or a survey has been used.
 - a. Fifty people at a shopping centre are asked to nominate the supermarket where they do most of their grocery shopping.
 - b. To determine the most popular new car on the road, 300 new-car buyers are asked what make and model they purchased.
 - c. To determine the most popular new car on the road, the make and model of every new registered car are recorded.
 - d. To test the life of a light bulb, every 100th bulb is tested.
5. For each of the following, recommend whether you would use a census or a survey to determine:
 - a. the most popular TV program on Sunday night at 8.30 pm
 - b. the number of 4-wheel-drive cars sold in a year
 - c. the number of cars travelling on a toll road each day
 - d. the percentage of defective SIM cards produced by a mobile phone manufacturing company.

Understanding

6. **WE2** Ten students were asked their opinions about the cause of congestion at the school's front gate. Analyse their responses below, suggest categories into which they could be classified and identify the most commonly stated reasons for the congestion.

1. The cars shouldn't come up the front driveway.
2. The front entrance is too small.
3. There should be another entrance.
4. The buses are the problem.
5. Bike riders should have a separate entrance.
6. Parents don't care where they park.
7. Kids just sit around there talking.
8. The buses all arrive at the same time.
9. The road is not wide enough.
10. There should be a traffic control officer there to direct the traffic.

7. Fifteen students were asked to identify the thing they liked least in their Maths class. Classify the responses below into appropriate categories, then identify the main reasons.

1. It's too hard.
2. There's too much homework.
3. I can't understand the teacher.
4. It's boring.
5. The boys are too distracting.
6. The teacher favours the boys.
7. It's too much work.
8. I can't concentrate for that length of time.
9. We do something new every lesson.
10. The teacher expects too much.
11. I don't like working in groups.
12. Our teacher is too strict.
13. Our class is too big.
14. The work is not interesting.
15. I don't like giving presentations to the class.

8. **WE3** Identify the issue with each of the following questions, then rewrite each so that the meaning is clear, unbiased and understandable.

- a. How much money do you earn?
- b. Do you exercise regularly?
- c. Which venue do you prefer — the SCG, the MCG, the Gabba or the WACA?
- d. Do you support the Premier's policy on wildlife preservation?
- e. How tall are you in feet and inches?
- f. Did you buy your sneakers for comfort and quality?

9. Rewrite the following open questions in closed format.

- a. How much pocket money do you get each week?
- b. How do you travel to school?
- c. What type of destination do you prefer for a holiday?



Reasoning

10. **WE5** Comment on the selection of the sample in the following cases.
- The choir members were interviewed regarding the opening hours of the school swimming pool.
 - Spectators at a football match were interviewed after the match to get their views on the Prime Minister's new health package.
 - Visitors arriving at Sydney's international airport were interviewed regarding the transport system from there into the city.
 - Twenty spectators at the Australian Open tennis championship were interviewed regarding the appropriate time of the year to hold the event.
11. Explain what 'open' and 'closed' questions are. Give an example of an open question, then rewrite your question in a closed format.
12. Discuss how you could use a standard deck of cards to select a random sample from a small population.



Problem solving

13. **WE4** As part of their quality control program, staff members at Perplexing Puzzles select 10 puzzles for testing from a batch of 50 coming off a production line. Use a random number generator to select the 10 puzzles to be tested.
14. An assignment requires you to research which social media platform is the most popular at your school.
- Determine the population being considered.
 - Discuss how you could use a census to gather information for the assignment.
 - Describe how you could select a sample from the population to survey for the assignment.
15. A survey was conducted at a school to see whether students preferred watching sports or playing sports. Eighteen people preferred watching sports and 16 preferred playing sports. If 30 people completed the survey, and 2 people don't like to watch or play sports, evaluate how many like to do both.

10.3 Primary and secondary data

LEARNING INTENTION

At the end of this subtopic you should be able to:

- understand the difference between primary and secondary data
- understand various methods of collecting primary data, including observation, measurement, and surveys
- understand the source and reliability of secondary data.

10.3.1 Primary data

- Primary data** are data that you have collected yourself.
- A variety of methods of collecting primary data is available. These include observation, measurement, survey, experiment or simulation.

Observation

- Observation involves recording the behavioural patterns of people, objects and events in a systematic manner.
- The data can be collected as a disguised observation (respondents are unaware they are being observed) or undisguised observation (the respondents are aware). Closed-circuit television (CCTV) cameras are an example of people knowing that their movements are being recorded but not always being aware of where the recording takes place.
- Observations can be in a natural environment (for example, in a food hall), or a contrived environment (for example, a food-tasting session for a food company).
- Mechanical devices (for example, video cameras, CCTV, counting devices across a road) can also be used.

Measurement

- Measurement involves using a measuring device to collect data.
- This generally involves conducting an experiment of some type.
 - The height of everyone in your class can be measured.
 - The mass of all newborn babies can be collected.
 - A pedometer can measure the number of steps the wearer takes.



Surveys

- Surveys are questionnaires designed to interview people. Often the questionnaire requires many rewrites to ensure it is clear and unbiased.
- The interview can be in person — face to face — or by telephone. The advantage of an in-person interview is that you are able to see the reactions of those you are interviewing and explain particular questions, if necessary.
- Until the beginning of the COVID-19 pandemic, email was the most frequently used interviewing tool; however, platforms such as WebEx and Zoom have become more popular as methods of communication. Some of the advantages and disadvantages of an email survey are listed as follows.
 - *Advantages:*
 - It can cover a large number of people or organisations.
 - Wide geographic coverage is possible.
 - It avoids embarrassment on the part of the respondent.
 - There is no interviewer bias.
 - The respondent has time to consider responses.
 - It is relatively cheap.
 - *Disadvantages:*
 - The questions have to be relatively simple.
 - The response rate is often quite low.
 - The reliability of the answers is questionable.
 - There is no control over who actually completes the questionnaire.
 - Questionnaires may be returned incomplete.



Experiment

- Generally, when conducting an experiment, the data collected is quantitative.
- Particular care should be taken to ensure the experiment is conducted in a manner that would produce similar results if repeated.
- Care must be taken with the recording of results.
- The results must be in a form that can readily be analysed.
- All results need to be recorded, including unusual or unexpected outcomes.



Simulation

- Experiments such as rolling a die, tossing a coin or drawing a card from a deck may be conducted to model real-life situations.
- Simulations occur in areas such as business, engineering, medicine and scientific research.
- Simulations are often used to imitate real-life situations that may be dangerous, impractical or too expensive to explore by other means.



WORKED EXAMPLE 6 Designing a simulation

It is widely believed that there is an equal chance of having a boy or girl with each birth. Although genetics and the history of births in a family may influence the sex of the child, ignore those factors in this question.

- a. Design an experiment to simulate the chance of giving birth to a boy or a girl.
- b. Describe how your experiment could be conducted to determine the number of children a couple should have, on average, to ensure they have offspring of both sexes.

THINK

- a. Use a device that can simulate two outcomes that are equally likely.
This could be a random number generator to generate two integers, say 0 (representing a boy) and 1 (representing a girl).
A fair coin could be tossed so that a Head represents a boy and a Tail represents a girl.

- b. 1. Describe how the experiment will be conducted.

WRITE

- a. A fair coin will be tossed, with a Head representing a boy (B) and a Tail representing a girl (G).

- b. The experiment will be conducted 50 times, and a record kept of each experiment.
For each experiment, the coin will be tossed until both sexes appear. This may mean that there could be, for example, 7 trials in an experiment (GGGGGB) before both sexes are represented.

2. Display the table of results.

The table below shows the results of the 50 experiments.

Exp. no.	Results	No. of trials	Exp. no.	Results	No. of trials
1	BG	2	26	GGGB	4
2	GGB	3	27	GGGGB	5
3	BG	2	28	GGGB	4
4	GGGGB	5	29	BG	2
5	BBBBBBG	7	30	BBBG	4
6	GGGB	4	31	BG	2
7	BBG	3	32	GB	2
8	BBG	3	33	GGGB	4
9	BBBG	5	34	BG	2
10	GB	2	35	GGGGGB	7
11	BG	2	36	BBBBBBG	7
12	GGGB	4	37	GB	2
13	BBG	3	38	BG	2
14	BBG	3	39	GGB	3
15	GB	2	40	GGGGB	5
16	BG	2	41	BBG	3
17	GGB	3	42	BBBBBBG	6
18	GB	2	43	GGB	3
19	GGB	3	44	GGB	3
20	BBBG	4	45	BBBG	4
21	BG	2	46	BBG	3
22	GB	2	47	GGGGGB	7
23	GGGGB	5	48	BG	2
24	BG	2	49	BBG	3
25	GGGGB	5	50	GGGGGB	6
				Total	175

This table shows that 175 trials were undertaken in 50 experiments where each experiment resulted in both sexes.

$$\begin{aligned} \text{Average number of children} &= \frac{175}{50} \\ &= 3.5 \end{aligned}$$

The average number of children a couple should have to reach the goal of having both sexes is 4.

3. Determine the average number of children required to produce offspring of both sexes.
4. Write a conclusion.

- Before collecting any primary data, it must be clear what data is to be collected.
- A decision must be made as to the method of collection.
- The advantages and disadvantages of the collection method must be acknowledged.
- The reason for the data collection should be clear from the outset.

WORKED EXAMPLE 7 Collecting data

You have been asked to obtain primary data to determine the methods of transport the students at your school use to travel to school. The data collected is meant to provide support for the student council's proposal for a school bus.

- State what data should be collected.
- Outline possible methods that could be used to collect the data.
- Decide which method you consider to be the best, and discuss its advantages and disadvantages.

THINK

- Outline the various forms of transport available to the students.
- Consider all the different ways of collecting the data.

1. Decide on the best option.

2. Discuss the advantages and disadvantages.

WRITE

- The modes of transport available to students at the school are car, bus, train, bicycle and walking.
- Several methods could be used to collect the data.
 - Stand at the school gate one morning and ask students as they arrive.
 - Design a questionnaire.
 - Ask students to write their mode of transport on a piece of paper and then place it in a collection tin.
- The first option (standing at the school gate) is time-consuming, and students could arrive at another entrance. The third option does not seem reliable, as some students may not comply, and other students may place multiple pieces of paper in the collection tin. The second option seems the best of the three.
Advantages of a questionnaire include:
 - There is a permanent record of responses.
 - It is not as time-consuming to distribute or collect.
 - Students can complete it at their leisure.Disadvantages of a questionnaire include:
 - Students may not return or complete it.
 - Printing copies could get expensive.

Note: This example does not represent the views of all those collecting such data. It merely serves to challenge students to explore and discuss available options.

- Sometimes the primary data required is not obvious at the outset of the investigation.
- For example, you are asked to investigate the claim:
Most students do not eat a proper breakfast before school.
What questions would you ask to prove or refute this claim?

10.3.2 Secondary data

- Secondary data** is data that has already been collected by someone else.
- The data can come from a variety of sources:
 - Books, journals, magazines, company reports
 - Online databases, broadcasts, videos
 - Government sources — the Australian Bureau of Statistics (ABS) provides a wealth of statistical data
 - General business sources — academic institutions, stockbroking firms, sporting clubs
 - Media — newspapers, TV reports
- Secondary data sources often provide data that would not be possible for an individual to collect.
- Data can be qualitative or quantitative — that is, categorical or numerical.

- The accuracy and reliability of data sometimes needs to be questioned, depending on its source.
- The age of the data should always be considered.
- It is important to learn the skills necessary to critically analyse secondary data.

WORKED EXAMPLE 8 Understanding data

Bigbite advertise the energy and fat content of some of the sandwiches on their menus.

- Determine the information that can be gained from this data.
- Bigbite advertise that they have a range of sandwiches with less than 6 grams of fat. Comment on this claim.
- This could be the starting point of a statistical investigation. Discuss how you could proceed from here.
- Investigations are not conducted simply for the sake of investigating. Suggest some aims for investigating further.

BIGBITE			
Bigbite fresh sandwiches	Energy (kJ)	Fat (g)	Sat fat (g)
Roasted vegetable	900	3.0	1.0
Ham	1100	6.0	1.4
Turkey	1140	4.8	1.7
BBQ beef	1150	5.0	1.5
Bigbite ribbon	1130	4.8	1.3
Turkey and ham	1250	4.5	1.5
BBQ chicken	1460	4.7	1.2
Chicken teriyaki	1110	4.0	1.0
Fresh dessert			
Fruit slice	200	<1	<1

Bigbite sandwiches
Regular sandwiches include white and/or wholemeal bread, salads and meat.
Nutritional value is changed by adding cheese or sauces.

THINK

- Look at the data to gain as much information as possible.

WRITE

- The data reveals the following information:
 - Higher energy content of a sandwich does not necessarily mean that the fat content is higher.
 - As the fat content of a sandwich increases, generally the saturated fat content also increases.
 - The addition of some types of protein (ham, turkey, beef, chicken) increases the energy content of the sandwich.
 - The data is only for those sandwiches on white or wholemeal bread with salads and meat.
 - The addition of condiments (sauces) or cheese will alter these figures.
 - A fruit slice has much less energy and fat than a sandwich.
- All the sandwiches displayed have less than 6 grams of fat, so Bigbite's claim is true. It must be remembered that the addition of cheese and sauce to these sandwiches would increase their fat content. Also, if the sandwich was on any bread other than white or wholemeal, the fat content could be higher than 6 grams.
- Conducting a web search for Bigbite's contact details, or to see whether more nutritional information is posted on their website, would be a good next step.

d. What are some interesting facts that could be revealed through a deeper investigation?

d. Suggested aims for investigating further could be:

- How much extra fat is added to a sandwich by the addition of cheese and/or sauce?
- What difference does a different type of bread make to the fat content of the sandwich?
- Which sandwich contains the highest fat content?
- What is the sugar content of the sandwiches?

DISCUSSION

Discuss some of the difficulties that you may come across with obtaining data from either primary or secondary sources.

Consider where the data would need to be collected from, the reliability of the data, what digital technologies might be needed and anything else that may influence the results.

COLLABORATIVE TASK: Simulations

Working in small groups, design an experiment to simulate the following situation.

A restaurant menu features 4 desserts that are assumed to be equally popular. How many dessert orders must be filled (on average) before the owner can be sure all types will have been ordered?

Carry out the experiment and discuss the results of the experiment with the class.

Discuss whether your answer would change if the menu features 6 desserts, all equally popular.



Resources

- **Worksheet** Topic 10 Workbook (worksheets, code puzzle and project) (ewbk-1941)
- **Interactivities** Individual pathway interactivity: Primary and secondary data (int-4452)
Primary and secondary data (int-3814)

Individual pathways

PRACTISE

1, 4, 8, 11

CONSOLIDATE

2, 5, 9, 12

MASTER

3, 6, 7, 10, 13

To answer questions online and to receive **immediate corrective feedback** and **fully worked solutions** for all questions, go to your learnON title at www.jacplus.com.au.

Fluency

1. **WE6** Devise an experiment to simulate each of the following situations and specify the device used to represent the outcomes.
 - a. A true/false test is used in which answers are randomly distributed.
 - b. A casino game is played, with outcomes grouped in colours of either red or black.
 - c. Breakfast cereal boxes are bought containing different types of plastic toys.
 - d. From a group of six people, one person is to be chosen as the group leader.
 - e. A choice is to be made between three main meals on a restaurant's menu, all of which are equally popular.
 - f. Five possible holiday destinations are offered by a travel agent; all destinations are equally available and equally priced.

2. **WE7** You have been asked to obtain primary data from students at your school to determine what internet access students have at home. The data collected will provide support for opening the computer room for student use at night.
 - a. Suggest what data should be collected.
 - b. Outline possible methods that could be used to collect this data.
 - c. Decide which method you consider to be the best option, and discuss its advantages and disadvantages.

3. **WE8** This label shows the nutritional information of Brand X rolled oats.

Nutrition Information			
Servings Per Package: 25		Serving Size 30 g	
	Per Serving 30 g	%DI* Per Serving	Per 100 g
Energy	486 kJ	6%	1620 kJ
Protein	4.3 g	9%	14.3 g
Fat - Total	2.8 g	4%	9.3 g
- Saturated	0.5 g	2%	1.7 g
- Trans	Less than 0.1 g	-	Less than 0.1 g
- Polyunsaturated	1.0 g	-	3.2 g
- Monounsaturated	1.3 g	-	4.4 g
Carbohydrate	16.8 g	5%	56.0 g
- Sugars	0.9 g	1%	3.0 g
Dietary Fibre	3.1 g	10%	10.4 g
Sodium	0.7 mg	0.1%	2 mg

* %DI = Percentage daily intake

- a. State the information gained from this data.
 - b. This could be the starting point of a statistical investigation. Discuss how you could proceed from here.
 - c. Suggest some aims for investigating further.
4. State which of the following methods could be used to collect primary data.
 Census, observation, newspaper article, journal, online response, DVD, interview, experiment, TV news report

Understanding

5. State which of the five methods below is the most appropriate to use to collect the following primary data.

Survey, observation, newspaper recordings, measurement, census

- Heights of trees along the footpaths of a tree-lined street
 - Number of buses that transport students to your school in the morning
 - Sunrise times during summer
 - Student opinion regarding length of lessons
6. Comment on this claim.

We surveyed 100 people to find out how often they eat chocolate.
Sixty of these people said they regularly eat chocolate.
We then measured the heights of all 100 people.
The conclusion:
eating chocolate makes you taller!

7. The following claim has been made regarding secondary data.

There's a lot more secondary than primary data. It's a lot cheaper and it's easier to acquire.

Comment on this statement.

Reasoning

8. Pizza King conducted a survey by asking their customers to compare 10 of their pizza varieties with those of their nearby competitors. After receiving and analysing the data, they released an advertising campaign with the headline 'Customers rate our pizzas as 25% better than the rest!'. The details in the small print revealed that this was based on the survey of their Hawaiian pizzas. Explain what was wrong with Pizza King's claim.
9. Addison, a prospective home buyer, wishes to find out the cost of a mortgage from financial institutions. She realises that there are a lot of lenders in the marketplace. Explain how she would collect the necessary information in the form of:
- primary data
 - secondary data
10. The local Bed Barn was having a sale on selected beds by Sealy and SleepMaker. Four of the beds on sale were:

Sealy Posturepremier	on sale for \$1499	a saving of \$1000
Sealy Posturepedic	on sale for \$2299	a saving of \$1600
SleepMaker Casablanca	on sale for \$1199	a saving of \$800
SleepMaker Umbria	on sale for \$2499	a saving of \$1800

The store claimed that all these beds had been discounted by at least 40%. Comment on whether this statement is true, supporting your comments with sound mathematical reasoning.

Problem solving

11. Hannah has two different data sets. Data set A contains the newborn baby weights of each student in her class after she surveyed each student. Data set B contains the average newborn baby weights for the last twenty years.
- Identify which data is primary data.
 - Identify which data is secondary data.
 - Explain how you determined which data was primary and which data was secondary.

12. Hamish is planning on running a stall at a fundraiser selling ice-cream. There are 1000 students in his school ranging from Year 7 to Year 12. There are five Year 8 classes, each with 25 students (boys and girls).

Hamish intends to ask a group of 10 students chosen at random from each of these five classes to select their favourite three ice-cream flavours. Hamish is confident that this random sampling method encompassing a total of 50 students should give him an accurate picture of the ice-cream preferences for the school. Is Hamish correct or is he facing a financial disaster? Explain your answer.



13. Kirsty, chief marketing manager of Farmco Cheeses, has decided to run a major TV advertising campaign.
- Suggest how she should choose a TV channel and time slot to run her advertisements.
 - Suggest how she should decide which demographic/age groups to target.
 - Discuss whether the answer to part b has any bearing on the answer to part a.

10.4 Organising and displaying data

LEARNING INTENTION

At the end of this subtopic you should be able to:

- organise data into a frequency table, using class intervals where necessary
- construct a histogram from a frequency table
- use technology to construct a histogram.



eles-4442

10.4.1 Examining data

- Once collected, data must be organised so that it can be displayed graphically and interpreted.
- When this has been done, any anomalies in the data will be highlighted.
- Anomalies could have occurred because of:
 - recording errors
 - unusual responses.
- Sometimes a decision is made to disregard these anomalies, which are regarded as **outliers**.
- Outliers can greatly affect the results of calculations, as you will see later in the topic.



Frequency tables

- Organising raw data into a frequency table is the first step in allowing us to see patterns (trends) in the data.
- A frequency table lists the values (or scores) of the variable and their frequencies (how often they occur).

WORKED EXAMPLE 9 Constructing a frequency table

In a suburb of 350 houses, a sample of 20 households was surveyed to determine the number of children living in them. The data were collected and recorded as follows:

0, 2, 3, 2, 1, 3, 5, 2, 0, 1, 2, 0, 2, 1, 2, 1, 2, 3, 1, 0

- Organise the data into a frequency table.
- Comment on the distribution of the data.
- Comment on the number of children per household in the suburb.

THINK

- Draw up a frequency table and complete the entries.

WRITE

a.

Children per household	Frequency
0	4
1	5
2	7
3	3
4	0
5	1

- Look at how the data are distributed.

- Does this sample seem to reflect the population characteristics?

b. The data value of 5 appears to be an outlier. This is probably not a recording error, but it is not typical of the number of children per household. Most households seem to have 1 or 2 children.

c. The sample is an appropriate size, and would probably reflect the characteristics of the population. It would be reasonably safe to say that most houses in the suburb contained 1 or 2 children.

- Sometimes data can take a large range of values (for example, age (0–100)) and listing all possible ages would be tedious. To solve this problem, we group the data into a small number of convenient intervals, called **class intervals**.
 - Class intervals should generally be the same size and be set so that each value belongs to one interval only.
 - Examples of class intervals are $0 < 5$, $5 < 10$, $10 < 15$ and so on; intervals represent the range of values that a particular group can take.
 - For example, the interval $0 < 5$ means numbers from 0 up to 5 (but not including 5) are contained within this group.

WORKED EXAMPLE 10 Constructing a frequency table with class intervals

A sample of 40 people was surveyed about the number of hours per week they spent watching TV. The results, rounded to the nearest hour, are listed below.

12, 18, 9, 17, 20, 7, 24, 16, 9, 27, 7, 16, 26, 15, 7, 28, 11, 20, 9, 11, 23, 19, 29, 12, 19, 12, 16, 21, 8, 4, 16, 20, 17, 10, 24, 21, 5, 13, 29, 26

- Organise the data into a frequency table using class intervals of $5 < 10$, $10 < 15$ and so on. Show the midpoint of each class interval.
- Comment on the distribution of the data.



THINK

- a. 1. Draw up a frequency table with three columns: class interval (hours of TV), midpoint and frequency.
 2. The midpoint is calculated by adding the two extremes of the class interval and dividing by 2. For example, the midpoint of the first class interval is $\frac{5 + 10}{2} = 7.5$.
 3. Systematically go through the list, determine how many times each score occurs and enter the information into the frequency column.
- b. Look at how the data is distributed.

WRITE

- a.
- | Hours of TV | Midpoint | Frequency |
|-------------|----------|-----------|
| 5– < 10 | 7.5 | 9 |
| 10– < 15 | 12.5 | 7 |
| 15– < 20 | 17.5 | 10 |
| 20– < 25 | 22.5 | 8 |
| 25– < 30 | 27.5 | |
- b. The TV viewing times are fairly evenly distributed, with the most frequent class interval being 15– < 20 hours per week.

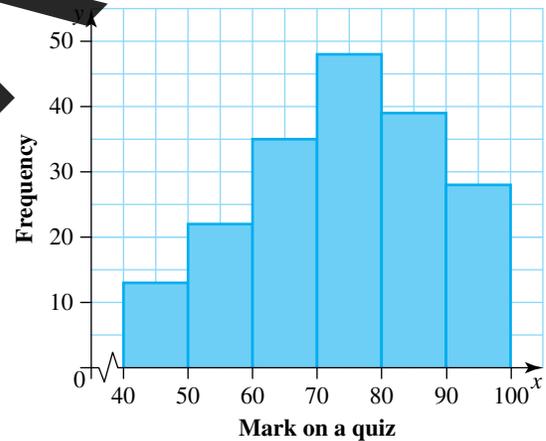


eles-4444

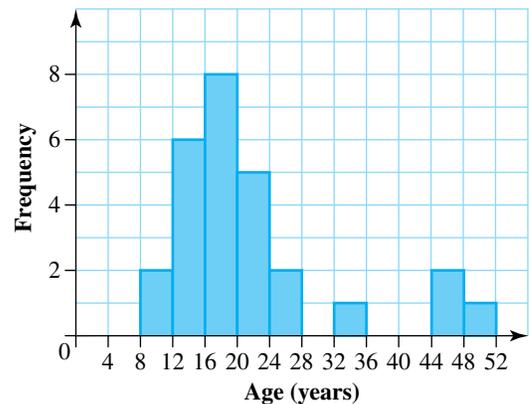
10.4.2 Histograms

- **Histograms** are used for displaying grouped discrete or continuous numerical data and can be used to highlight trends and distributions.
- Histograms display data that has been summarised in a frequency table.
- A histogram has the following characteristics:
 1. The vertical axis (y-axis) is used to represent the frequency of each item.
 2. No gaps are left between columns.
 3. A space measuring a half-column width is sometimes placed between the vertical axis and the first column of the histogram if the first bar does not start at zero.
- Before constructing a histogram, identify the smallest and largest values for both axes to help choose an appropriate scale.
- Both axes must be labelled. The vertical axis is labelled 'Frequency'.
- When presenting grouped data graphically, we generally label the horizontal axis (score) with the class interval.

Grouped discrete data



Continuous numerical data



WORKED EXAMPLE 11 Constructing a histogram from grouped data

Consider the grouped frequency table created in Worked example 10.

- Display the data as a histogram.
- Comment on the shape of the graph.

Hours of TV	Frequency
5– < 10	9
10– < 15	7
15– < 20	10
20– < 25	8
25– < 30	6

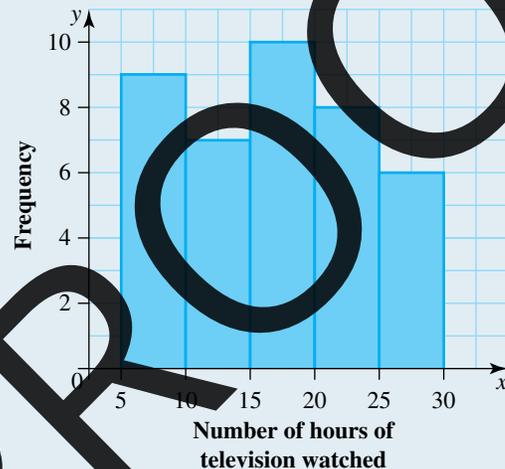
THINK

- Rule a set of axes on graph paper. Give the graph a title. Label the horizontal axis 'Number of hours of television watched' and the vertical axis 'Frequency'.
 - Leaving a $\frac{1}{2}$ unit space at the beginning, draw the first column so that it starts at 5 and ends at the beginning of the next interval, which is 10. The height of this first column should be 9. Repeat this technique for the other scores.

WRITE/DRAW

a.

Histogram of hours of television watched



- Look at how the data is distributed.

- The number of hours of TV watching is fairly consistent throughout the week. A maximum number of people watch about 15 to 20 hours per week.

- The next example illustrates how to deal with a large amount of data, organise it into class intervals, then produce a histogram.

WORKED EXAMPLE 12 Organising and displaying data

The following data are the results of testing the lives (in hours) of 100 torch batteries.

20, 31, 42, 49, 46, 36, 42, 25, 28, 37, 48, 49, 45, 35, 25, 42, 30, 23, 25, 26,
 29, 31, 46, 25, 40, 30, 31, 49, 38, 41, 23, 46, 29, 38, 22, 26, 31, 33, 34, 32,
 41, 23, 29, 30, 29, 28, 48, 49, 31, 49, 48, 37, 38, 47, 25, 43, 38, 48, 37, 20,
 38, 22, 21, 33, 35, 27, 38, 31, 22, 28, 20, 30, 41, 49, 41, 32, 43, 28, 21, 27,
 20, 39, 40, 27, 26, 36, 36, 41, 46, 28, 32, 33, 25, 31, 33, 25, 36, 41, 28, 33

- Choose a suitable class interval for the given data and present the results in a frequency distribution table.
- Draw a histogram of the data.

THINK

- a. 1. To choose a suitable size for the class intervals, calculate the range. To determine the range, subtract the smallest value from the largest.
2. Divide the results obtained for the range by 5 and round to the nearest whole number. *Note:* A class interval of 5 hours will result in 6 groups.
3. Draw a frequency table and list the class intervals in the first column, beginning with the smallest value. *Note:* The class interval 20– < 25 includes hours ranging from and including 20 to less than 25.
4. Systematically go through the data and determine the frequency of each class interval.
5. Calculate the total of the frequency column.

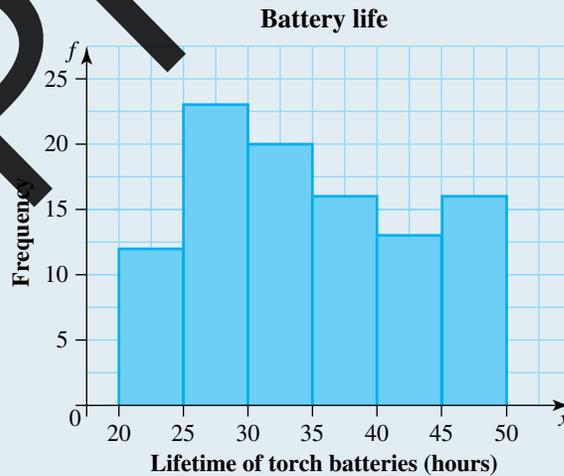
WRITE/DRAW

a. Range = largest value – smallest value
= 49 – 20
= 29

Number of class intervals: $\frac{29}{5} = 5.8$
= 6

Lifetime (hours)	Tally	Frequency (f)
20– < 25		12
25– < 30		23
30– < 35		20
35– < 40		16
40– < 45		13
45– < 50		16
Total		100

- b. 1. Rule and label a set of axes on graph paper. Give the graph a title.
2. Add scales to the horizontal and vertical axes. *Note:* Leave a half interval at the beginning and end of the horizontal axis.
3. Draw in the first column so that it starts at 20 and finishes at 25 and reaches a vertical height of 12 units.
4. Repeat step 3 for each of the other scores.



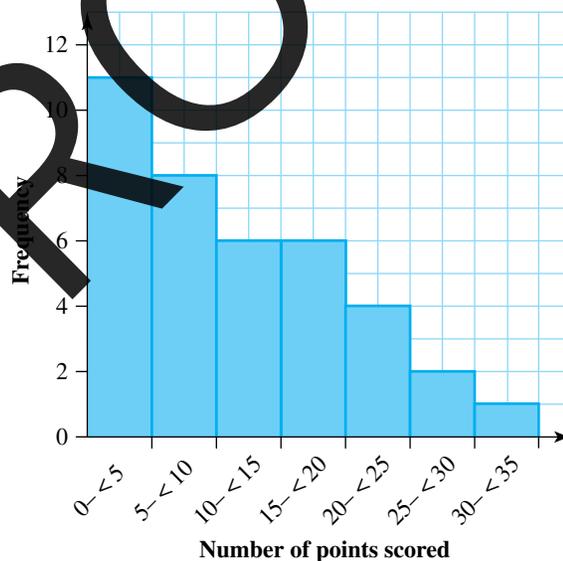
10.4.3 Using a spreadsheet to draw a histogram

eles-4445

- Spreadsheets such as Excel can also be used to easily tabulate and graph data.
- Enter the data values or class intervals into the first column and frequencies into the second column, as shown.

	A	B
1	Number of points scored	Frequency
2	0– < 5	11
3	5– < 10	8
4	10– < 15	6
5	15– < 20	6
6	20– < 25	4
7	25– < 30	2
8	30– < 35	1

- To construct a histogram, follow the steps outlined below:
 1. Highlight all the cells in your table containing data.
 2. Click on the **Insert** tab at the top of the screen.
 3. Select the **Clustered Column** graph from the Charts section.
 4. Next to the column graph click on the + symbol and select **Axis Titles**.
 5. Change the vertical Axis Title to 'Frequency'.
 6. Change the horizontal Axis Title to the name of the data (in this case 'Number of points scored').
 7. Change the column graph to a histogram by right-clicking on any column and selecting **Format Data Series**. Change the Gap Width to 0%.



on Resources

-  **eWorkbook** Topic 10 Workbook (worksheets, code puzzle and project) (ewbk-1941)
-  **Digital document** SKILLSHEET Presenting data in a frequency table (doc-6989)
-  **Interactivities** Individual pathway interactivity: Organising and displaying data (int-4453)
 - Frequency tables (int-3816)
 - Column graphs (int-3817)

Individual pathways

PRACTISE

1, 3, 6, 9, 12

CONSOLIDATE

2, 5, 8, 10, 13

MASTER

4, 7, 11, 14

To answer questions online and to receive **immediate corrective feedback** and **fully worked solutions** for all questions, go to your learnON title at www.jacplus.com.au.

Fluency

1. **WE9** In a suburb of roughly 1500 houses, a random sample of 40 households was surveyed to calculate the number of children living in each. The data were collected and recorded as follows.

0, 3, 2, 4, 1, 2, 3, 2, 2, 2, 2, 1, 3, 4, 5, 2, 3, 1, 1, 1,
0, 0, 2, 3, 4, 1, 3, 4, 2, 2, 0, 1, 2, 3, 2, 0, 2, 4, 5, 1

- Organise the data into a frequency table.
- Comment on the distribution of the data.
- Comment on the number of children per household in the suburb.



2. A quality control officer selected 25 boxes of smart watches at random from a production line. She tested every single smart watch and displayed the number of defective smart watches in each box as follows:

1, 3, 2, 5, 2, 2, 1, 5, 2, 1, 2, 4, 3, 0, 5, 3, 2, 1, 3, 2, 1, 3, 4, 2, 1

- Comment on the sample.
- Organise the data into a frequency table.
- Comment on the distribution of the data.
- Comment on the population of smart watches.



3. **WE11** This table shows the number of hours of sport played per week by a group of Year 8 students.

Score (hours of sport played)	Frequency (<i>f</i>)
1 < 2	3
2 < 3	8
3 < 4	10
4 < 5	12
5 < 6	16
6 < 7	8
7 < 8	7
Total	64



- Draw a histogram to display the data.
- Comment on the shape of the graph.
- Discuss whether you feel this sample reflects the sporting habits of Year 8 students generally.

4. A block of houses in a suburb was surveyed to determine the size of each house (in m^2). The results are shown in the following table.

Size of house (m^2)	Frequency
100– < 150	13
150– < 200	18
200– < 250	19
250– < 300	17
300– < 350	14
350– < 400	11
Total	92

- Draw a histogram to display the data.
 - Comment on the shape of the graph.
 - Discuss whether you feel this sample reflects the size of the houses in the suburb.
5. Forty people joined a weight-loss program. Their mass (in kg) was recorded at the beginning of the program and is shown in the frequency table.

Class interval	Frequency
60– < 70	2
70– < 80	5
80– < 90	9
90– < 100	12
100– < 110	7
110– < 120	3
120– < 130	2
Total	40

- Draw a histogram to display the data.
- Comment on the shape of the graph.
- Discuss whether you feel this sample reflects the mass of people in the community.

Understanding

6. **WE10, 12** Forty people in a shopping centre were asked about the number of hours per week they spent watching TV. The result of the survey is shown as follows.

10, 13, 7, 12, 16, 11, 6, 14, 6, 11, 5, 14, 12, 8, 27, 17, 13, 8, 14, 10,
13, 7, 15, 10, 16, 8, 18, 14, 21, 28, 9, 12, 11, 13, 9, 13, 29, 5, 24, 11

- Organise the data into class intervals of 5– < 10 hours, and so on, and draw up a frequency table.
 - Draw a histogram to display the data.
 - Comment on the shape of the graph.
 - Discuss whether you feel this sample reflects the TV-viewing habits of the community.
7. The number of hours of sleep during school weeknights for a Year 8 class are recorded below.

6, 9, 7, 8, 7, $8\frac{1}{2}$, $6\frac{1}{2}$, 8, $7\frac{1}{2}$, $7\frac{1}{2}$, 8, $8\frac{1}{2}$, $6\frac{1}{2}$, 8, 8, 7, $7\frac{1}{2}$, 8, 9, 8

- Organise the data into suitable class intervals and display it as a frequency table.
- Display the data as a histogram.
- Comment on the sleeping habits of the Year 8 students.
- Discuss whether you feel these sample results reflect those of Year 8 students generally.

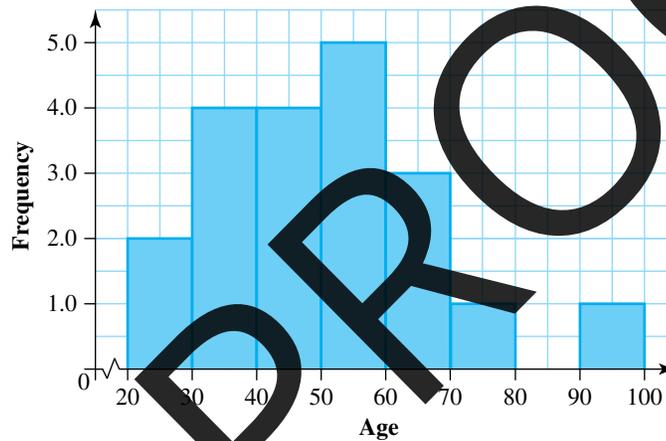
8. The amount of pocket money (in dollars) available to a random sample of 13-year-olds each week was found to be as shown below.

10, 15, 5, 4, 8, 10, 4, 15, 5, 6, 10, 6, 5, 10, 8, 10, 5, 10, 10, 6

- Organise the data into class intervals of $0- < 5$, $5- < 10$ dollars, and so on, and display it as a frequency table.
- Display the data as a histogram.
- Comment on the shape of the histogram.
- Discuss whether you feel these sample results reflect those of 13-year-olds generally.

Reasoning

- Show that the midpoint for the interval $12.5- < 13.2$ is 12.85.
- The following histogram shows ages of patients treated by a doctor during a shift.



Complete a frequency table for the histogram.

- The following data gives the results of testing the lives (in hours) of 100 torch batteries.

25, 36, 30, 34, 21, 40, 36, 46, 29, 38, 20, 41, 34, 45, 25, 40, 31, 39, 24, 45, 27, 44, 23, 35, 47, 49, 20, 37, 43, 26, 35, 28, 48, 30, 20, 36, 41, 26, 32, 42, 21, 31, 45, 42, 26, 37, 33, 24, 45, 38, 36, 43, 21, 34, 38, 35, 28, 41, 30, 22, 29, 32, 39, 25, 44, 21, 35, 38, 41, 35, 30, 23, 37, 43, 33, 54, 28, 39, 22, 31, 35, 42, 38, 27, 36, 46, 28, 34, 37, 29, 24, 30, 39, 44, 31, 24, 36, 28, 47, 21

- Choose a suitable class interval for the given data, and present the results in a frequency distribution table.
- Draw a histogram of the data.
- Comment on the trends shown by the histogram.
- Discuss whether you feel these results reflect those of the battery population.



Problem solving

12. A building company recorded the number of weekends during which their tradespeople needed to work over the course of one year, as shown in the table.

Class interval	Frequency
1–4	4
5–8	6
9–12	13
13–16	11
17–20	5
21–24	7
25–28	10

- Identify how many tradespeople work at the company.
- Determine the most common number of weekends worked.
- Is it possible to determine the maximum number of weekends worked by a tradesperson? Explain.



13. The following data were collected on the number of times people go to the cinema per month.

4, 5, 7, 9, 1, 2, 5, 2, 4, 8, 3, 6, 2, 3, 8, 1, 1, 4,
5, 3, 3, 6, 1, 2, 7, 1, 3, 2, 2, 4, 10, 0, 1, 3, 4, 6

- Organise the data into class intervals of 0–2, 3–5 etc. and display it as a frequency table.
 - Draw a histogram to represent the data.
 - Determine how many people go to the cinema fewer than three times a month.
 - Determine how many people go to the cinema at least three times a month.
 - Is it reasonable to draw conclusions about the whole population based on this sample? Give reasons for your answer.
14. A random sample of 30 students in Year 9 undertook a survey to investigate the heights of Year 9 students. These were their measured heights (in cm).

146, 163, 156, 168, 159, 170, 152, 174, 156, 163, 157, 161, 178, 151, 148,
167, 162, 157, 166, 154, 150, 166, 160, 155, 164, 157, 171, 168, 158, 162

- Organise the data into class intervals of 145– < 150 cm and so on, and display it as a frequency distribution table.
- Draw a histogram displaying the data.
Reorganise the class intervals into 145– < 148 cm and so on, and construct a new frequency distribution table.
- Draw a new histogram displaying the data in part c.
- Comment on the similarities and differences between the two histograms.

10.5 Measures of centre and spread

LEARNING INTENTIONS

At the end of this subtopic you should be able to:

- determine the mean, median and mode of a set of data
- determine the range of a set of data
- identify possible outliers in a set of data and understand how they affect the mean, median and range.



eles-4446

10.5.1 Mean

- The **measures of centre** or **measures of location** give some idea of the average or middle of the data set.
- The two types of measures of centre used in interpreting data are the **mean** and **median**.
- The mean is the average of a set of scores, and is denoted by the symbol \bar{x} (pronounced *x bar*).
- The mean can only be calculated for a set of numerical data.
- The value of the mean of a set of data is often not one of the given scores.

Determining the mean

To calculate the mean (or average) of a set of data, use the following formula.

$$\bar{x} = \frac{\text{sum of the data values (or scores)}}{\text{total number of data values}}$$

WORKED EXAMPLE 13 Calculating the mean

Jan's basketball scores were 18, 24, 20, 22, 14 and 12. Calculate his mean score, correct to 1 decimal place.

THINK

1. Calculate the sum of the basketball scores.
 2. Count the number of basketball scores.
 3. Define the rule for the mean.
 4. Substitute the known values into the rule.
 5. Evaluate, rounding to 1 decimal place.
- Note:* Jan's typical (or average) score per game is 18.3 points.

WRITE

$$\begin{aligned}\text{Sum of scores} &= 18 + 24 + 20 + 22 + 14 + 12 \\ &= 110\end{aligned}$$

$$\text{Total number of scores} = 6$$

$$\text{Mean} = \frac{\text{sum of the data values}}{\text{total number of data values}}$$

$$\begin{aligned}\bar{x} &= \frac{110}{6} \\ &= 18.333\ 33\ \dots \\ &= 18.3\end{aligned}$$

- Sometimes calculations need to be performed from a frequency distribution table.
- Calculating the mean from a frequency table requires a different process than that used for raw data.

Determining the mean from a frequency table

To calculate the mean (or average) from a frequency table, use the following formula.

$$\bar{x} = \frac{\text{total of (frequency} \times \text{score) column}}{\text{total of frequency column}}$$

WORKED EXAMPLE 14 Calculating the mean from a frequency table

Calculate the mean of the frequency distribution data given below correct to 1 decimal place.

Score (x)	Frequency (f)
1	3
2	2
3	4
4	0
5	5

THINK

- Copy and complete the frequency table and include an extra column called frequency \times score ($f \times x$).
- Enter the information into the third column. The score of 1 occurred 3 times. Therefore, $f \times x = 3 \times 1 = 3$. The score of 2 occurred 2 times. Therefore, $f \times x = 2 \times 2 = 4$. Continue this process for each pair of data.
- Determine the total of the 'Frequency' column. This shows how many scores there are altogether.
- Determine the total of the 'Frequency \times score' column. This shows the sum of the values of all the scores.
- Define the rule for the mean.
- Substitute the known values into the rule.
- Evaluate the answer to 1 decimal place.
Note: The typical (or average) value of the set of data is 3.1.

WRITE

Score (x)	Frequency (f)	Frequency \times score ($f \times x$)
1	3	$3 \times 1 = 3$
2	2	$2 \times 2 = 4$
3	4	$4 \times 3 = 12$
4	0	$0 \times 4 = 0$
5	5	$5 \times 5 = 25$
Total	14	44

$$\text{Mean} = \frac{\text{total of (frequency} \times \text{score) column}}{\text{total of frequency column}}$$

$$\begin{aligned}\bar{x} &= \frac{44}{14} \\ &= 3.142857\dots \\ &= 3.1\end{aligned}$$

10.5.2 Median

- The median is the middle value if the data values are placed in numerical (ascending) order.
- The median can only be calculated for a set of numerical data.

Determining the median

The following formula determines the *position* of the median value of a set of scores in numerical order.

$$\text{location of median} = \left(\frac{n+1}{2}\right)\text{th score, in a set of } n \text{ scores}$$

Note: This formula does not determine the median value. It simply locates its position in the data set.

- For sets of data containing an odd number of scores, the median will be one of the actual scores; for sets with an even number of scores, the median will be positioned halfway between the two middle scores.

WORKED EXAMPLE 15 Determining the median

Determine the median of each of the following sets of scores.

a. 10, 8, 11, 5, 17

b. 9, 3, 2, 6, 3, 5, 9, 8

THINK

a. 1. Arrange the values in numerical (ascending) order.

2. Select the middle value.

Note: There is an odd number of scores: 5. Hence, the third value is the middle number or

median. Alternatively, the rule $\frac{n+1}{2}$, where $n = 5$, gives the position of the median. The location of the median is $\left(\frac{5+1}{2} = 3\right)$; that is,

the 3rd score.

3. Write the answer.

b. 1. Arrange the values in ascending order.

2. Select the two middle values.

Note: There is an even number of scores: 8. Hence, the fourth and fifth values are the middle numbers, or median. Again the rule $\frac{n+1}{2}$ could be used to locate the position of the median.

3. Obtain the average of the two middle values (the fourth and fifth values).

4. Write the answer.

WRITE

a. 5, 8, 10, 11, 17

5, 8, 10, 11, 17

The median of the scores is 10.

b. 2, 3, 3, 5, 6, 8, 9, 9

Location of median = $\frac{n+1}{2}$
 $= \frac{8+1}{2}$
 $= \frac{9}{2}$
 $= 4.5\text{th value}$
(i.e. between the fourth and fifth values)

$$\begin{aligned}\text{Median} &= \frac{5+6}{2} \\ &= \frac{11}{2} \\ &= 5\frac{1}{2} \text{ (or } 5.5\text{)}\end{aligned}$$

The median of the scores is $5\frac{1}{2}$ or 5.5.

10.5.3 Mode

eles-4448

- The mode is the most common score or the score with the highest frequency in a set of data. It is *not* considered to be a measure of centre.
- The mode measures the clustering of scores.
- Some sets of scores have more than one mode or no mode at all. There is no mode when all values occur an equal number of times.
- The mode can be calculated for both numerical and categorical data.

Determining the mode

- The mode is the most common score or the score with the highest frequency.
- Some data sets have one unique mode, more than one mode or no mode at all.

WORKED EXAMPLE 16 Determining the mode

Determine the mode of each of the following sets of scores.

a. 5, 7, 9, 8, 5, 8, 5, 6

b. 10, 8, 11, 5, 17

c. 9, 3, 2, 6, 3, 5, 9, 8

THINK

a. 1. Look at the set of data and circle any values that have been repeated.

2. Choose the values that have been repeated the most.

3. Write the answer.

b. 1. Look at the set of data and circle any values that have been repeated.

2. Answer the question.

Note: No mode is not the same as a mode that equals 0.

c. 1. Look at the set of data and circle any values that have been repeated.

2. Choose the values that have been repeated the most.

3. Write the answer.

WRITE

a. 5, 7, 9, 8, 5, 8, 5, 6

The number 5 occurs three times.

The mode for the given set of values is 5.

b. 10, 8, 11, 5, 17

No values have been repeated.

The following set of data has no mode, since none of the scores has the highest frequency. Each number occurs only once.

c. 9, 3, 2, 6, 3, 5, 9, 8

The number 3 occurs twice. The number 9 occurs twice.

The modes for the given set of values are 3 and 9.

WORKED EXAMPLE 17 Calculating measures of centre

The data from a survey asking people how many times per week they purchased takeaway coffee from a cafe are shown below.

2, 9, 11, 8, 5, 5, 5, 8, 7, 4, 5, 3

Use the data to calculate each of the following.

- The mean number of coffees purchased per week
- The median number of coffees purchased per week
- The modal number of coffees purchased per week

THINK

a. 1. To calculate the mean, add all the values in the data set and divide by the total number of data values. There are 12 values in the data set.

2. Write the answer.

b. 1. The median is the value in the middle position. There are 12 values in the data set, so the middle position is between the 6th and 7th values.

2. Arrange the data set in order from lowest to highest.
The 6th value is 5.
The 7th value is 5.

3. Write the answer.

c. The mode is the most common value in the data set. The most common value is 5.

WRITE

$$\begin{aligned} \text{a. } \bar{x} &= \frac{\text{sum of all the values}}{\text{total number of values}} \\ &= \frac{2+9+11+8+5+5+5+8+7+4+5+3}{12} \\ &= \frac{72}{12} \\ &= 6 \end{aligned}$$

The mean number of coffees purchased per week is 6.

$$\begin{aligned} \text{b. Location of median} &= \frac{n+1}{2} \\ &= \frac{12+1}{2} \\ &= \frac{13}{2} \\ &= 6.5 \end{aligned}$$

2, 3, 4, 5, 5, 5, 7, 8, 8, 9, 11

$$\begin{aligned} \text{Median} &= \frac{5+5}{2} \\ &= \frac{10}{2} \\ &= 5 \end{aligned}$$

The median number of coffees purchased per week is 5.

c. The modal number of coffees purchased per week is 5.

- If the shape of a distribution for a set of data is symmetrical, then the mean and median values will be the same. This implies that the average value and the middle score will be the same.



eles-4449

10.5.4 Measures of spread

- In analysing a set of scores, it is helpful to see not only how the scores tend to cluster, or how the middle of the set looks, but also how they spread or scatter.
- For example, two classes may have the same average mark, but the spread of scores may differ considerably.
- The **range** of a set of scores is the difference between the highest and lowest scores.
- The range is a **measure of spread**.

Determining the range

To determine the range of a set of data, use the following formula.

$$\text{range} = \text{highest score} - \text{lowest score}$$

- The range can only be calculated for a set of numerical data.

WORKED EXAMPLE 18 Calculating the range

Calculate the range of the following sets of data.

a. 7, 3, 5, 2, 1, 6, 9, 8

b.

Score (x)	Frequency (f)
7	1
8	3
9	5
10	2

THINK

- a.
 1. Obtain the highest and lowest values.
 2. Define the range.
 3. Substitute the known values into the rule.
 4. Evaluate.
 5. Write the answer.
- b.
 1. Obtain the highest and lowest values.
Note: Consider the values (scores) only, not the frequencies.
 2. Define the range.
 3. Substitute the known values into the rule.
 4. Evaluate.
 5. Write the answer.

WRITE

- a. Highest value = 9
Lowest value = 1
Range = highest value – lowest value
 $= 9 - 1$
 $= 8$
The set of values has a range of 8.
- b. Highest value = 10
Lowest value = 7
Range = highest value – lowest value
 $= 10 - 7$
 $= 3$
The frequency distribution table data has a range of 3.

- Although the range identifies both the lowest and highest scores, it does not provide information on how the data is spread out between those values.
- In most cases, the spread of data between the lowest and highest scores is not uniform.

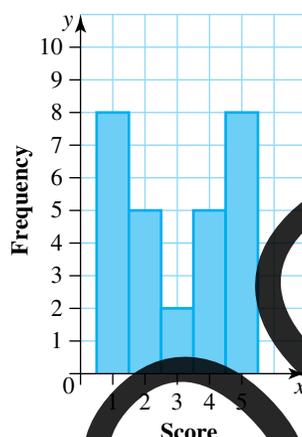
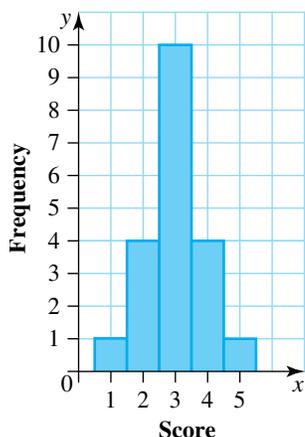
COLLABORATIVE TASK: Analyse this!

Equipment: Data collected from the survey in *Collaborative Task: Designing a survey*, from subtopic 10.2.

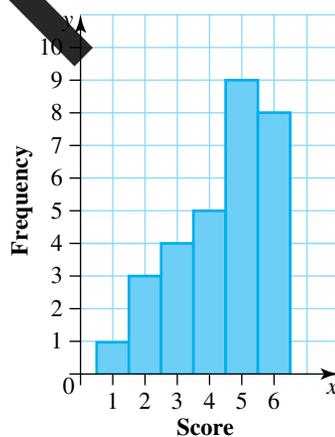
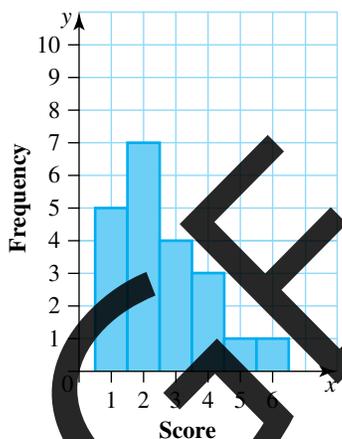
1.
 - a. For your collected data, calculate the mean, median and mode for each of the questions you asked.
 - b. Are these statistics appropriate for the type of data you collected in each question? Think about what these values mean for your data and whether the values you are achieving are appropriate for the type of data you have collected.
2. Choose an appropriate visual representation for the data you collected for each question. What do you need to take into consideration before selecting the visual representation?
3.
 - a. Select the data from one question to represent as grouped data using class intervals.
 - b. Calculate the mean, median and modal class for the group data.
 - c. Compare the values of mean, median and mode for the grouped data with those of the ungrouped data. What do you notice? Suggest a reason for anything you noticed.
4. As a class, discuss any similarities and differences you found between the statistics for your grouped and ungrouped data.

10.5.5 Clusters, gaps and outliers

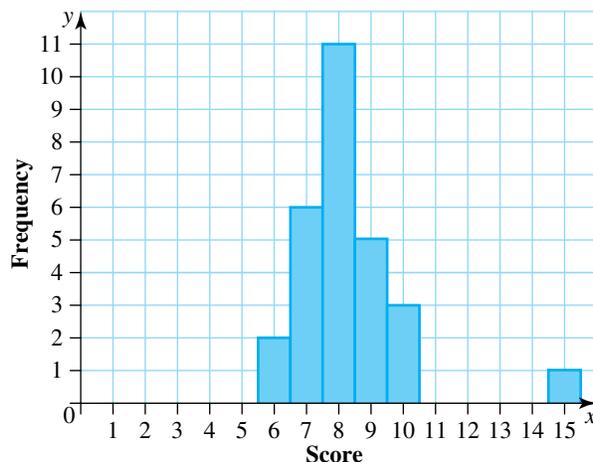
- Clusters, gaps and outliers in the data set can be seen in histograms.
- A cluster is a grouping of data points that are close together.
- Consider the following two histograms.

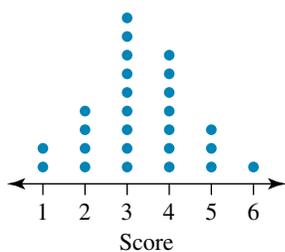


- In the first histogram, the mean, median and mode of the data set is 3 and the data is clustered around the mean.
- In the second histogram, the data is not clustered and there are two modes — one at either end of the distribution. The mean and median of this data set are also 3, but the modes are now 1 and 5.
- Data can also be clustered at either the lower end or the upper end of a distribution, as shown in the following histograms.



- Data values that do not follow the general pattern of the distribution could be classified as **outliers**.
- Looking at this histogram, we can see that data value 15 does not follow the general pattern of the distribution and is possibly an outlier.
- Clusters, gaps and outliers can also be identified in other data representations, including stem-and-leaf plots and dot plots.





Key: 1 | 2 = 12

Stem	Leaf
1	2 4 6
2	0 1 3 7
3	2 3 4 4 6 7 8
4	0 1 5
5	
6	
7	8

The effect of outliers on measures of centre and spread

- The presence of one or more outliers may have a considerable effect on the measures of centre and spread of a particular set of data.

WORKED EXAMPLE 19 Determining the effect of an outlier

A netball team scored the following points in 10 games:

17, 23, 31, 19, 50, 29, 16, 23, 30, 32

- Calculate the mean, median, mode and range of the team scores.
- The following Saturday, the regular goal shooter was ill, and Lauren, who plays in a higher division, was asked to play. The team's score for that game was 200. Recalculate the mean, median, mode and range after 11 games.
- Comment on the similarities and differences between the two sets of summary statistics.

THINK

- Rearrange the scores in numerical order.
- Calculate the mean.
- Calculate the median.
- Calculate the mode.
- Calculate the range.
- Write the answer.

WRITE

a. 16, 17, 19, 23, 23, 29, 30, 31, 32, 50

$$\bar{x} = \frac{16 + 17 + 19 + 23 + 23 + 29 + 30 + 31 + 32 + 50}{10} = 27$$

$$\text{Median} = \frac{23 + 29}{2} = 26$$

23 appears the most times.

$$\text{Range} = 50 - 16 = 34$$

$$\text{Mean} = 27$$

$$\text{Median} = 26$$

$$\text{Mode} = 23$$

$$\text{Range} = 34$$

- Rearrange the 11 scores in numerical order.
- Calculate the mean.
- Calculate the median.
- Calculate the mode.
- Calculate the range.
- Write the answers.

b. 16, 17, 19, 23, 23, 29, 30, 31, 32, 50, 200

$$\frac{16 + 17 + 19 + 23 + 23 + 29 + 30 + 31 + 32 + 50 + 200}{11} = 43$$

The middle value of the data set is 29.

23 appears the most times.

$$200 - 16 = 184$$

$$\text{Mean} = 43$$

$$\text{Median} = 29$$

$$\text{Mode} = 23$$

$$\text{Range} = 184$$

- c. Compare the two sets of summary statistics.

The inclusion of an extreme value or outlier has dramatically increased the mean and the range of the data, marginally increased the median and left the mode unchanged.

Note: The important point to learn from Worked example 19 is that when a set of data includes extreme values, the mean may not be truly representative of the data.

Which measure of centre is most useful?

- It is important to know which measure of centre will be most useful in a given situation
 - The mean is appropriate when no extreme values or outliers distort the picture.
 - The median is appropriate when outliers are present.
 - The mode is appropriate when the most common result is significant.

Resources

-  **eWorkbook** Topic 10 Workbook (worksheets, code puzzle and project) (ewbk-1941)
-  **Digital documents**
 - SkillSHEET Finding the mean of ungrouped data (doc-6991)
 - SkillSHEET Arranging a set of data in ascending order (doc-6992)
 - SkillSHEET Finding the median (doc-6993)
 - SkillSHEET Finding the middle score for data arranged in a dot plot (doc-6994)
 - SkillSHEET Finding the score in a data set that occurs most frequently (doc-6995)
-  **Video eLesson** Mean and median (eles-1905)
-  **Interactivities**
 - Individual pathway interactivity: Measures of centre and spread (int-4454)
 - Mean (int-3818)
 - Median (int-3819)
 - Mode (int-3820)
 - Range (int-3822)
 - Outliers (int-3821)

Exercise 10.5 Measures of centre and spread

learnON

Individual pathways

■ PRACTISE

1, 2, 6, 11, 12, 13, 17, 21, 23, 26, 27, 30

■ CONSOLIDATE

3, 4, 8, 9, 15, 18, 19, 22, 24, 28, 31

■ MASTER

5, 7, 10, 14, 16, 20, 25, 29, 32, 33

To answer questions online and to receive **immediate corrective feedback** and **fully worked solutions** for all questions, go to your learnON title at www.jacplus.com.au.

Fluency

1.  Caroline's basketball scores were 28, 25, 29, 30, 27 and 22. Calculate her mean score correct to 1 decimal place.
2. Calculate the mean (average) of each set of the following scores. Give the answers correct to 2 decimal places.
 - a. 1, 2, 3, 4, 7, 9
 - b. 2, 7, 8, 10, 6, 9, 11, 4, 9
 - c. 3, 27, 14, 0, 2, 104, 36, 19, 77, 81
 - d. 4, 8.4, 6.6, 7.0, 7.5, 8.0, 6.9

3. Francesca's soccer team has the following goals record this season:
2, 0, 1, 3, 1, 2, 4, 0, 2, 3

- State the total number of goals it has scored.
- State the number of games the team has played.
- Calculate the team's average score.

4. **MC** Frisco's athletics coach timed 5 consecutive 200-metre training runs. He recorded times of 25.1, 23.9, 24.8, 24.5 and 27.3 seconds. His mean 200-metre time (in seconds) is:

- A. 24.60 B. 25.20 C. 25.12
D. 25.42 E. 26.12

5. Two Year 8 groups did the same Mathematics test. Their results out of 10 were:

Group A: 5, 8, 7, 9, 6, 7, 8, 5, 4, 2

Group B: 5, 6, 4, 5, 9, 7, 8, 8, 9, 7

Determine which group had the highest mean.

6. **WE14** Calculate the mean of this frequency distribution, correct to 2 decimal places.

Score (x)	Frequency (f)
1	4
2	3
3	6
4	1
5	0

7. Calculate the mean of this frequency distribution, correct to 2 decimal places.

Score (x)	Frequency (f)
6	2
7	8
8	3
9	6
10	2

8. **WE15a** Calculate the median of the following scores.

a. 5, 5, 7, 12, 13

b. 28, 13, 17, 21, 18, 17, 14

9. **WE15b** Calculate the median of each of the following sets of scores.

a. 52, 46, 52, 48, 52, 48

b. 1.5, 1.7, 2.0, 1.8, 1.5, 1.7, 1.8, 1.9

10. **WE16** For each set of scores in questions 8 and 9, state the mode.



Questions 11 and 12 refer to the following set of scores.

1, 1, 1, 4, 4, 5, 5, 6, 3, 3, 7, 6, 5, 4, 6, 2, 1, 8

11. **MC** The median of the given scores is:

- A. 1
- B. 4.5
- C. 4
- D. 5
- E. 8

12. **MC** The mode of the given scores is:

- A. 5
- B. 6
- C. 4
- D. 3
- E. 1

13. **WE18a** Calculate the range of the following scores.

- a. 5, 5, 7, 12, 13
- b. 28, 13, 17, 21, 18, 17, 14
- c. 2, 52, 46, 52, 48, 52, 48
- d. 4, 1.5, 1.7, 2.0, 1.8, 1.5, 1.7, 1.8, 1.9

14. **WE18b** Determine the range of the following sets of data.

a.

Score (x)	Frequency (f)
6	1
7	5
8	10
9	7
10	3

b.

Score (x)	Frequency (f)
1	7
2	9
3	6
4	8
5	10
6	10

c.

Score (x)	Frequency (f)
5	1
10	5
15	10
20	7

d.

Score (x)	Frequency (f)
110	2
111	2
112	2
113	3
114	3

15. Determine the range of each of the following data sets.

a. Key: 1 | 8 = 18

Stem	Leaf
1	1 2 7 8 9
2	2 8
3	1 3 7 9
4	0 1 2 6

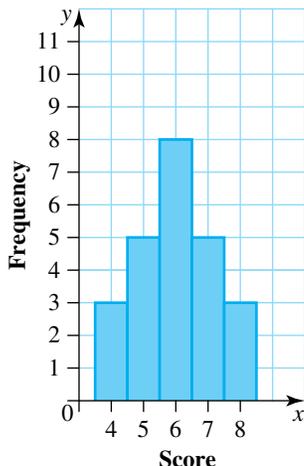
b. Key: 24 | 7 = 247

Stem	Leaf
24	2 7
25	2 4 6 6 8
26	0 1 3 5 9
28	5 6 6 8

c. Key: 17 | 4 = 174

Stem	Leaf
15	6 2 4
16	8 6 1 3 9
17	0 2 1 8 6 7 3 4
18	4 1 5 2 7 1

16. Calculate the mean, median, mode and range of the data shown in the following histogram.



Understanding

17. A third Year 8 group had the following results in the same test as in question 5.

Group C: 5, 7, 8, 4, 6, 8, 5, 9, 8

- a. Calculate the average score of this group, correct to 1 decimal place.
- b. Determine the score that a tenth student (who was originally absent) would need to achieve to bring this group's average to 7.

18. A survey of the number of occupants in each house in a street gave the following data:

2, 5, 1, 6, 2, 3, 2, 4, 1, 2, 0, 2, 3, 2, 4, 5, 4, 2, 3, 4

Prepare a frequency distribution table with an $f \times x$ column and use it to calculate the mean number of people per household.

19. The mean of 5 scores is 7.2

- a. Calculate the sum of the scores.
- b. If four of the scores are 9, 8, 7 and 5, determine the fifth.

20. Over 10 matches, a soccer team scored the following number of goals:

2, 3, 1, 0, 4, 5, 2, 3, 3, 4

- a. Identify the most common number of goals scored.
 - b. Identify the median number of goals scored.
- In this case, determine whether the mode or the median give a score that shows a typical performance.



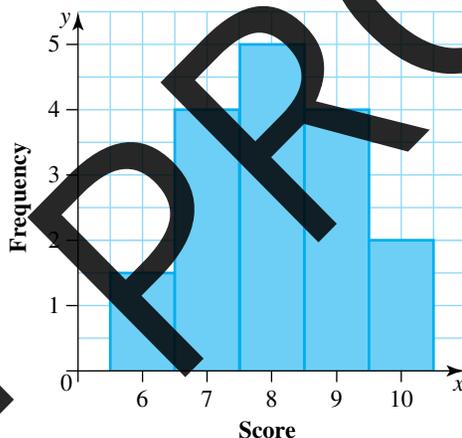
21. **WE17** The following scores represent the number of muesli bars sold in a school canteen each day over two weeks:

54, 64, 51, 58, 56, 59, 10, 34, 48, 56

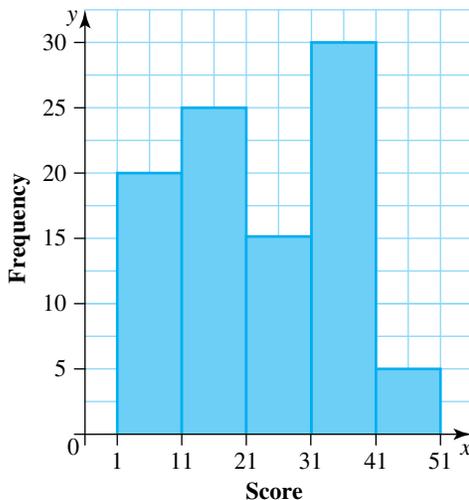
- Calculate the mean.
 - Calculate the median.
 - Calculate the mode.
 - Of the mean, median and mode, explain which best represents a typical day's sales at the school canteen.
22. A small business pays the following annual wages (in thousands of dollars) to its employees:

18, 18, 18, 18, 26, 26, 26, 40, 80

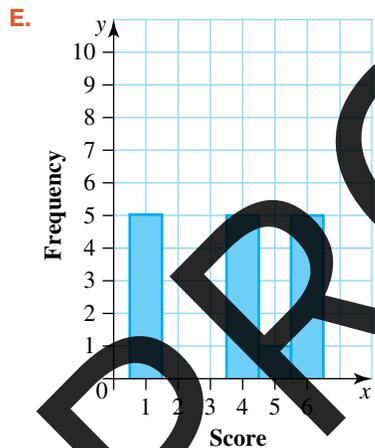
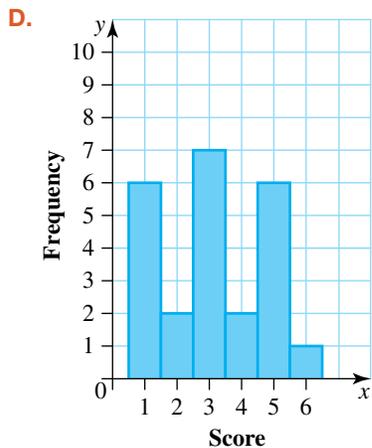
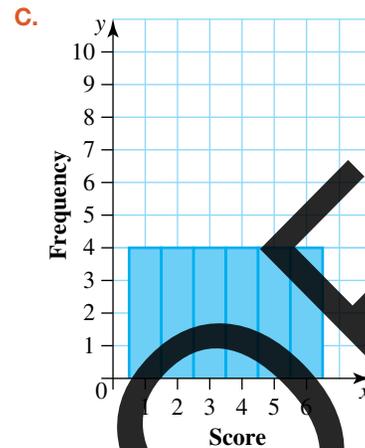
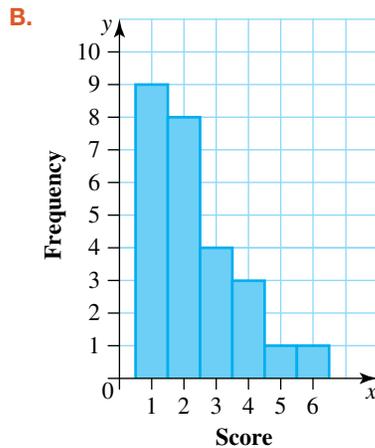
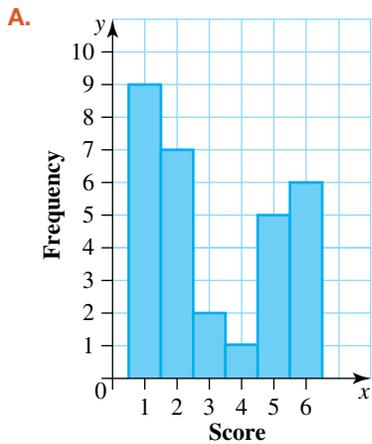
- Identify the mode of the distribution.
 - Identify is the median wage.
 - Calculate is the mean wage.
 - Explain which measure you would expect the employees' union to use in wage negotiations.
 - Discuss which measure the boss might use in such negotiations.
23. The following histogram shows the distribution of a set of scores.



- Identify the mode.
 - Determine whether there are any outliers in the data set. If so, calculate their value(s).
24. Consider the following distribution. Determine the modal class.



25. **MC** Select which of the following data sets shows data that is clustered.



26. **WE19** A rugby team scored the following numbers of points in 10 games:

24, 18, 33, 29, 22, 16, 38, 30, 26, 30

- Calculate the mean, median, mode and range of the team scores.
- The following Saturday, the team played a side who had lost their previous 35 games and who were on the bottom of the ladder. The team's score for that game was 86. Recalculate the mean, median, mode and range after 11 games.
- Comment on the similarities and differences between the two sets of summary statistics.



Reasoning

27. These data show the number of hours Year 8 students used a computer in a particular week:

5, 3, 6, 7, 3, 5, 2, 5, 2, 3, 6, 7

- Calculate the mean, median and mode.
- Comment on the value of the mean compared with the median.
- Comment on the value of the mean compared with the mode.
- Explain which is the best indicator of the centre of the data set.



28. Determine which measure of centre is most appropriate to use in the following situations. Explain your answers.
- Analysing property values in different suburbs of a capital city
 - Determining the average shoe size sold at a department store
 - Determining the average number of tries scored over a season of rugby league

29. The following scores represent data from an online survey asking about the average number of hours students spent exercising each week:

3, 5, 1, 4, 0, 8, 23, 4, 2, 0, 2, 6

- Identify the potential error in the data set.
- Explain whether this data value could have been a genuine outlier.



Problem solving

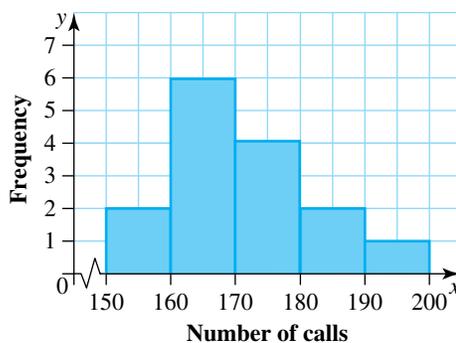
30. Calculate the mean number of books presented in this frequency table, using the midpoint of each interval as the x -value for the interval.

Number of books (x)	Frequency (f)
1–15	3
16–30	9
31–45	8
46–60	11
61–75	10
76–90	14
91–105	15
106–120	18



31. Identify the mean, median and mode in the following paragraph:
It was an amazing game of cricket today. The winning team hit more sixes than any other number of runs. This meant that, even though the middle value of runs per over was 3, the sixes brought the average up to about five runs per over. What an incredible game!
32. The mean of 5 different test scores is 15. Evaluate the largest and smallest possible test scores, given that the median is 12. All test scores are whole numbers. Justify your answer.
33. Evaluate the mean number of calls made on mobile phones in the month shown in the graph, using the midpoint of each interval to represent the number of phone calls per month.

Monthly mobile phone calls



10.6 Analysing data

LEARNING INTENTION

At the end of this subtopic you should be able to:

- analyse a data set using summary statistics
- make predictions about a population from a sample.

10.6.1 Analysing data sets

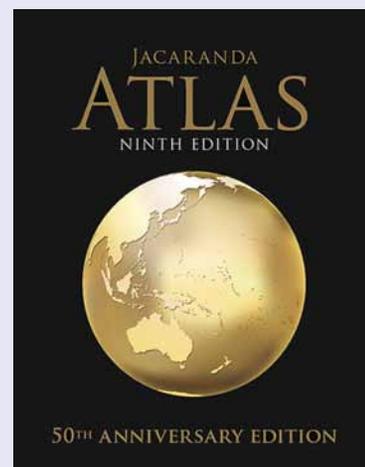
eles-4450

- The mean, median, mode and range are collectively known as **summary statistics** and play an important role in analysing data.
- To analyse a data set:
 - calculate the measures of centre — mean and median
 - determine the mode
 - calculate the spread — range
 - construct frequency tables and histograms.
- Remember:
 - the data comes from surveys of samples of the population
 - what each statistical measure gives.

Statistical measures	Definition and purpose
Mode	The most common score or category. It tells us nothing about the rest of the data. Data may have no mode, one mode or more than one mode.
Median	The score in the exact middle of the values placed in numerical order. It provides information about the centre of the distribution. It tells us nothing about the rest of the data. It is unaffected by exceptionally large or small scores (outliers).
Mean	The typical (or average) score expected. It can be calculated as the sum of all the scores divided by the number of scores and is affected by exceptionally large or small scores (outliers).
Range	The difference between the highest score and the lowest score. It shows how far the scores are spread apart. It is particularly useful when combined with the mean or the median. It is affected by outliers.

ACTIVITY: Analysing data from other subjects

Look for data from one of your other school texts. For example, your Geography text or atlas may contain information about sustainable food production or population statistics. Calculate and compare sample summary statistics for the data you have found.



WORKED EXAMPLE 20 Identifying the statistical measure

Explain which statistical measure is referred to in these statements.

- The majority of people surveyed prefer Activ-8 sports drink.
- The ages of fans at the Rolling Stones concert varied from 8 to 80.
- The average Australian family has 2.5 children.

THINK

- Write the statement and highlight the keyword(s).
 - Relate the highlighted word to one of the statistical measures.
 - Answer the question.
- Write the statement and highlight the keyword(s).
 - Relate the highlighted word to one of the statistical measures.
 - Answer the question.
- Write the statement and highlight the keyword(s).
 - Relate the highlighted word to one of the statistical measures.
 - Answer the question.

WRITE

- The **majority** of people surveyed prefer Activ-8 sports drink.
Majority implies most, which refers to the mode.
This statement refers to the mode.
- The ages of fans at the Rolling Stones concert **varied** from 8 to 80.
The statement refers to the range of fans' ages at the concert.
This statement refers to the range.
- The **average** Australian family has 2.5 children.
The statement deals with surveying the population (census) and finding out how many children are in each family.
This statement refers to the mean.



eles-4451

10.6.2 Using a sample to predict the properties of a population

- Once survey data has been collated and analysed, the data set can be used to predict the characteristics of the population from which it was taken. Consider this example.

WORKED EXAMPLE 21 Making predictions

The 153 students in Year 8 all sat for a 10-question multiple-choice practice test for an upcoming exam. A random sample of the results of 42 of the students gave this distribution.

Score (x)	Frequency (f)
1	2
2	3
3	6
4	7
5	11
6	8
7	4
8	0
9	0
10	1

- Calculate the mean mark, correct to 1 decimal place.
- Determine the median mark.
- Give the modal mark.
- Determine which measure of centre best represents the data.
- Comment on any prediction about the properties of the population from this sample.

THINK

- Add a third column called $f \times x$. Multiply the frequency by its corresponding mark in each row to complete the column.
 - Determine the totals of the frequency column and the $f \times x$ column.

WRITE

a.

Mark (x)	Frequency (f)	$f \times x$
1	2	2
2	3	6
3	6	18
4	7	28
5	11	55
6	8	48
7	4	28
8	0	0
9	0	0
10	1	10
Total	42	195

- Define the rule for the mean.
- Substitute the known values into the rule and evaluate, giving your answer correct to 1 decimal place.

$$\begin{aligned} \text{Mean} &= \frac{\text{total of } (f \times x) \text{ column}}{\text{total of frequency column}} \\ &= \frac{195}{42} \\ &= 4.6 \end{aligned}$$

- The median is the score in the middle; that is, the $\left(\frac{42+1}{2}\right)$ th score — the average of the 21st and 22nd score. Add a cumulative frequency column to the original frequency distribution. (Find the total frequency at that point for each mark.) Look in the cumulative frequency column to see where the 21st and 22nd scores lie.

b.

Mark (x)	Frequency (f)	Cumulative frequency
1	2	2
2	3	$2 + 3 = 5$
3	6	$5 + 6 = 11$
4	7	$11 + 7 = 18$
5	11	$18 + 11 = 29$
6	8	$29 + 8 = 37$
7	4	$37 + 4 = 41$
8	0	$41 + 0 = 41$
9	0	$41 + 0 = 41$
10	1	$41 + 1 = 42$
Total	42	42

The 21st and 22nd scores are both 5. The median is 5.

- c. The modal mark is the one that occurs most frequently. Look for the one with the highest frequency.
- d. Compare the results for the mean, median and mode. Look for similarities and differences.
- e. Consider whether these results from the sample would reflect those of the population.
- c. The mode is 5.
- d. The mean is 4.6, and the median is 5. It seems that either of these measures would be appropriate to use as a measure of centre of the data. However, check the mark of 10 as it could be a possible outlier. When the mark of 10 is disregarded, the mean is calculated to be $185 \div 41 = 4.5$. Since the value of the mean did not change significantly after removing the mark of 10, we can be safe in concluding that the mark of 10 is not an outlier. So, the mean or median could be used as a measure of centre of the data.
- e. It seems likely that these results would reflect those of the whole population. The sample is random and of sufficient size. The one perfect score of 10 indicates that there would be a few students with full marks, and at least half the students passed the test.

- It is important to note that summary statistics may vary from sample to sample even though they are taken from the same population.
- For example, if you collect data on heights of students and one sample consisted only of boys and a second sample only of girls, the statistical measures would vary significantly.

Using a spreadsheet to calculate summary statistics

- Spreadsheets such as Excel can calculate statistical measures.
- Enter the data values into a column as shown in column B, rows 2 to 9, in the spreadsheet shown.
- To calculate the mean, use the formula ‘=AVERAGE(’ and then select all of the cells containing your data. Close the brackets and the mean will be calculated.
Type ‘=AVERAGE(B2:B9)’ into cell B11, then press ENTER.
- To calculate the median, use the formula ‘=MEDIAN(’ and then select all of the cells containing your data. Close the brackets and the median will be calculated.
Type ‘=MEDIAN(B2:B9)’ into cell B12, then press ENTER.
- To calculate the mode, use the formula ‘=MODE(’ and select all of the cells containing your data. Close the brackets and the mode will be calculated.

Note: If there is more than one mode, this method will only display one of the modes, so double check the data set.

- Type ‘=MODE(B2:B9)’ into cell B13, then press ENTER.
- To calculate the range, use the formula ‘=MAX(’ and then select all of the cells containing your data. Close the brackets and type ‘-MIN(’ and again select all of the cells containing your data. Close the brackets and the range will be calculated.
Type ‘=MAX(B2:B9)-MIN(B2:B9)’ into cell B14, then press ENTER.

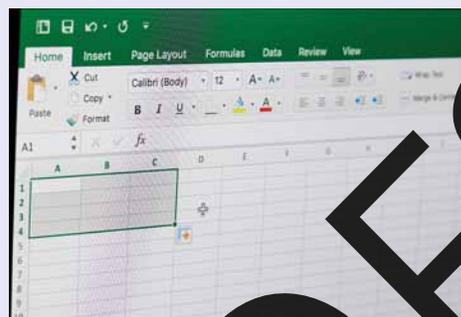
	A	B
1		Scores
2		5
3		10
4		7
5		5
6		8
7		12
8		6
9		11
10		
11	Mean	8
12	Median	7.5
13	Mode	5
14	Range	7

ACTIVITY: Using a spreadsheet to calculate and compare summary statistics

Pick a topic to use to investigate the population of students at your school — for example, the ages of their parents or guardians. Take three different samples from your population.

Enter your collected data into a spreadsheet and calculate the summary statistics for each of your different samples.

Compare the summary statistics for your three samples and comment on the similarities and differences.



Resources

 **eWorkbook** Topic 10 Workbook (worksheets, code puzzle and project) (ewbk1941)

 **Interactivity** Individual pathway interactivity: Analysing data (int-4456)

Exercise 10.6 Analysing data

learnON

Individual pathways

PRACTISE

1, 2, 4, 6, 11, 14

CONSOLIDATE

3, 5, 8, 12, 15

MASTER

7, 9, 10, 13, 16

To answer questions online and to receive **immediate corrective feedback** and **fully worked solutions** for all questions, go to your learnON title at www.jacplus.com.au.

Fluency

1. **WE20** Explain which statistical measure is referred to in these statements.

- There was a 15°C temperature variation during the day.
- Most often you have to pay \$79.95 for those sports shoes.
- The average Australian worker earns about \$1659 per week.
- A middle-income family consisting of 2 adults and 2 children earns about \$116 600 per annum.



2. **WE21** This frequency table shows the results of a random sample of 15 students (from a class of 30) who sat for a 10-question multiple-choice test.

- Calculate the mean mark.
- Determine the median mark.
- Give the modal mark.
- Which measure of centre best represents the data?
- Comment on any prediction of properties of the population from this sample.

Score (x)	Frequency (f)
4	1
5	2
6	5
7	4
8	3
Total	15

3. Consider the following frequency distribution tables.

a.

Score (x)	Frequency (f)
1	4
2	3
3	2
4	1
5	0

b.

Score (x)	Frequency (f)
6	2
7	8
8	3
9	4
10	2

For each one:

- calculate the mean score to 1 decimal place
- determine the median score
- identify the modal score
- indicate which measure of centre best describes the distribution.

4. Consider the following stem plots.

a. Key: 110 = 10

Stem	Leaf
1	0 2
2	1 3 3 5
3	
4	4

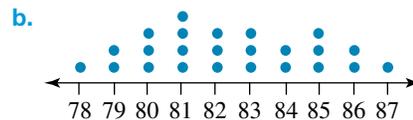
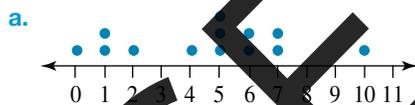
b. Key: 1010 = 100

Stem	Leaf
10	0
11	0 2 2 2
12	0 4 6 6
13	3

For each one:

- calculate the mean score to 1 decimal place
- determine the median score
- identify the modal score
- indicate which measure of centre best describes the distribution.

5. Consider the following dot plots.



For each one:

- calculate the mean score to 1 decimal place
- determine the median score
- identify the modal score
- indicate which measure of centre best describes the distribution.

Understanding

6. Three different samples looked at the average yearly incomes of NSW households. The figures given are in thousands of dollars (for example, \$97 000) and have been rounded to the nearest whole number.

Sample 1: 97, 135, 52, 106, 189, 158, 70, 81, 122, 69

Sample 2: 102, 131, 85, 204, 77, 85, 114, 90, 111, 126

Sample 3: 66, 89, 110, 90, 173, 77, 129, 166, 256, 98

- Calculate the mean, median, mode and range for each sample.
- Compare the summary statistics for the three different samples.

7. A survey of the number of people living in each house in a street produced the following data:

2, 5, 1, 6, 2, 3, 2, 1, 4, 3, 4, 3, 1, 2, 2, 0, 2, 4

- Display the data as a frequency table and determine the average (mean) number of people per household, correct to 1 decimal place.
- Draw a dot plot of the data and use it to determine the median number per household.
- Identify the modal number per household.
- State which of the measures would be most useful to:
 - a real-estate agent renting out houses
 - a government population survey
 - a mobile ice-cream vendor.



8. The contents of 20 packets of matches were counted after random selection. The following numbers were obtained:

138, 139, 139, 141, 137, 140, 137, 141, 139, 142, 140, 141, 141, 139, 141, 138, 139, 140, 141, 138

- Construct a frequency distribution table for the data.
- Determine the mode, median and mean of the distribution. Give answers correct to 1 decimal place where necessary.
- Comment on which of the three measures best supports the manufacturer's claim that there are 140 matches per box.



9. A class of 26 students had a median mark of 54 in Mathematics; however, no-one actually obtained this result.

- Explain how this is possible.
- Explain how many students must have scored below 54.

10. A soccer team had averaged 2.6 goals per match after 5 matches. After their sixth match, the average had dropped to 2.5. How many goals did they score in that latest match?



Reasoning

11. A tyre manufacturer selects 48 tyres at random from the production line for testing. The total distance travelled during the safe life of each tyre is shown in the table.

Distance in km ($\times 1000$)	46	50	52	56	78	82
Number of tyres	4	12	16	10	4	2

- Calculate the mean, median and mode.
 - Discuss which measure best describes 'average' tyre life. Explain your answer.
 - Recalculate the mean with the 6 longest-lasting tyres removed. By how much is it lowered?
 - If you selected a tyre at random, determine the distance it would be most likely to last.
 - In a production run of 10 000 tyres, determine how many could be expected to last for a maximum of 50 000 km.
 - As the manufacturer, explain for what distance you would be prepared to guarantee your tyres.
12. Read the following paragraph and explain what statistics are represented and what they mean.

It's been an exciting day at the races today. There was a record fast time of 38 seconds, and also a record slow time of 4 minutes and 52 seconds. We had an unbelievable number of people who ran the race in exactly 1 minute. Despite this, the average time was well over 2 minutes, due to the injury of a few runners.

13. If you take more than one sample from the same population, explain why the summary statistics will vary from sample to sample.

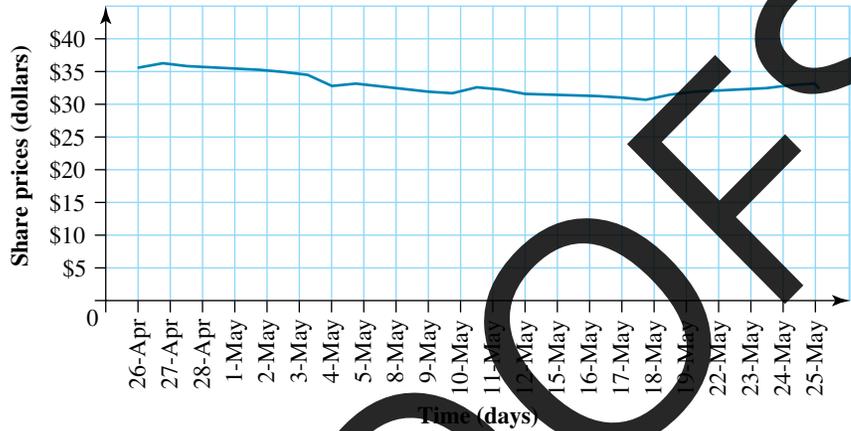
Problem solving

14. The following graph displays the movement of the price of BankSave shares over a 30-day period.

The closing price of the shares after 22 days of trading, rounded to the nearest dollar, were:

35, 36, 36, 35, 35, 34, 33, 33, 32, 32, 33, 32, 32, 31, 31, 31, 32, 32, 33, 33, 33, 33

Using the rounded amounts, calculate the mean (correct to 2 decimal places), median, mode and range of the share prices.



15. At a preview cinema session, the ages of the viewers were recorded and displayed in a stem-and-leaf plot. For this data, evaluate the:

- a. range
- b. mean
- c. median
- d. mode.

Key: 1 | 6 = 16 years

Stem	Leaf
1	5 6 7 7 8 9 9
2	1 2 4 8 8
3	0 1 1 1 5
4	2 3
5	3

16. The number of goals a netballer scored in the 12 games of a season was as follows:

1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 8, 12

A local newspaper reporter asked the netballer what her average was for the season.

- a. Explain which measure of centre (mean, median or mode) the netballer should give the reporter as her ‘average’ so that the value of the average is as high as possible.
- b. Explain which measure of centre you would choose to best describe the ‘average’ number of goals the netballer scored each game.



10.7 Review

10.7.1 Topic summary

Collecting data

- Data can be collected via:
 - observation
 - survey
 - experiment.
- A census is a survey of the population.
- Due to time and cost, samples are often surveyed.
- Samples should be randomly selected and the size = $\sqrt{\text{population size}}$.

Organising data

- Data can be displayed in many ways, such as:
 - frequency tables
 - histograms
 - spreadsheets.
- Where the range of values is large, the data may be grouped in class intervals, such as 5 or 10.
- Outliers (data points that differ significantly from the others) may be excluded from the analysis.

REPRESENTING AND INTERPRETING DATA

Measures of spread

- The range is a measure of spread. It is the difference between the highest and lowest values in the data set, including outliers.
$$\text{range} = \text{highest value} - \text{lowest value}$$
- The range is greatly affected by outliers.

Primary and secondary data

- Primary data is data you have collected.
- Methods of collecting primary data include observation, measurement, survey, experiment and simulation.
- Secondary data is data that has been collected by someone else.
- Sources of secondary data include magazines, journals, videos, television and websites.

Measures of centre

- The measures of centre are the mean and median.
- The mean, \bar{x} , is the average of the data set.
$$\bar{x} = \frac{\text{sum of data values}}{\text{total number of data values}}$$
- Or, if the data is from a frequency table:
$$\bar{x} = \frac{\text{total of (frequency} \times \text{score) column}}{\text{total of frequency column}}$$
- The median is the middle value of the data set.
$$\text{location of median} = \left(\frac{n+1}{2}\right)\text{th value}$$
- The mode is the most common value or value with the highest frequency.
- The mean is affected by outliers, whereas the median is not.



10.7.2 Success criteria

Tick a column to indicate that you have completed the subtopic and how well you think you have understood it using the traffic light system.

(**Green:** I understand; **Yellow:** I can do it with help; **Red:** I do not understand)

Subtopic	Success criteria			
10.2	I understand the difference between a sample and a population.			
	I understand the various methods of collecting data, such as a survey, census or questionnaire.			
	I can select an appropriate sample and know whether it is biased or not.			
10.3	I understand the difference between primary and secondary data.			
	I understand various methods of collecting primary data, including observation, measurement, and surveys.			
	I understand the source and reliability of secondary data.			
10.4	I can organise data into a frequency table, using class intervals where necessary.			
	I can construct a histogram from a frequency table.			
	I can use technology to construct a histogram.			
10.5	I can determine the mean, median and mode of a set of data.			
	I can determine the range of a set of data.			
	I can identify possible outliers in a set of data and I understand how they affect the mean, median and range.			
10.6	I can analyse a data set using summary statistics.			
	I can make predictions about a population from a sample.			

10.7.3 Project

Analysing the English language

The English alphabet contains 26 letters that combine to form words. Have you ever wondered why some letters appear more often than others? Scrabble® is a game that allows players to form interlocking words in a crossword style. The words are formed using lettered tiles that carry a numerical value. Players compete against each other to form words of the highest score until all the tiles have been used.



The following table displays the letter distribution and the value of each letter tile in the game of Scrabble.

Letter	Number of tiles	Letter score	Letter	Number of tiles	Letter score	Letter	Number of tiles	Letter score	Letter	Number of tiles	Letter score
A	9	1	B	2	3	C	2	3	D	4	2
E	12	1	F	2	4	G	3	2	H	2	4
I	9	1	J	1	8	K	1	5	L	4	1
M	2	3	N	6	1	O	8	1	P	2	3
Q	1	10	R	6	1	S	4	1	T	6	1
U	4	1	V	2	4	W	2	4	X	1	8
Y	2	4	Z	1	10	Blank	2	0			

1. State which letters are most common in Scrabble and which are least common.
2. What do you notice about the relationship between the number of tiles used for each letter and their letter score?

This passage about newspapers and magazines is taken from an English textbook.

Who reads news?

Newspapers and magazines are produced for different categories of readers. These are known as the 'target audience'. Particular groups of readers buy certain types of newspapers and magazines because these publications have content which interests them as a reader. Magazines and newspapers cover different types of news; for example, a local paper contains information on local issues whereas a magazine like *Movie* has all the latest news on films and film stars. General newspapers like *The Age* and *The Australian* try to appeal to a broader audience by including sections on various topics such as sport, business and property.

3. Complete a frequency table for the distribution of letters in this passage.
4. How do the results from this frequency table compare with the Scrabble game's frequency table?

Knowing which letters in the English language are most common can be very helpful when we are trying to solve coded messages. The following paragraph is written in code.

Tnlmktbtl ehvtmxw bg max lhnmaxkg axfbliaxkx tgv bl ftwx ni hy lbq lmtmxl tgv mph mxkkbmkhbxl. Max vtibmt ny Tnlmktbtl bl Vtguxkkt.

5. Study the coded message carefully. Use an appropriate method to decode the message.
6. Explain the strategies you used to decode the message.

Online Resources

-  **eWorkbook** Topic 10 Workbook (worksheets, code puzzle and project) (ewbk-1941)
-  **Interactivities** Crossword (int-2760)
Sudoku puzzle (int-3191)

To answer questions online and to receive **immediate corrective feedback** and **fully worked solutions** for all questions, go to your learnON title at www.jacplus.com.au.

Fluency

1. For each of the following statistical investigations, state whether a census or a survey has been used.
 - a. The average price of petrol in Sydney was estimated by averaging the price at 40 petrol stations.
 - b. The Australian Bureau of Statistics has every household in Australia complete an online questionnaire or information form every five years.
 - c. The performance of a cricketer is measured by looking at his performance in every match he has played.
 - d. Public opinion on an issue is sought by a telephone poll of 2000 homes.

2. **MC** Identify which of the following is an example of a census.
 - A. A newspaper conducts an opinion poll of 2000 people.
 - B. A product survey of 1000 homes is carried out to determine what brand of washing powder is used.
 - C. Every 200th jar of Vegemite is tested to see if it is the correct mass.
 - D. A federal election is held.
 - E. At a shopping centre, 500 people are questioned regarding the parking facilities at the centre.

3. **MC** Identify which of the following is an example of a random sample.
 - A. The first 50 students who arrive at school take a survey.
 - B. Fifty students' names are drawn from a hat, and those drawn take the survey.
 - C. Ten students from each year level at a school are asked to complete a survey.
 - D. One class at a school is asked to complete a survey.
 - E. Those who catch the bus to school are asked to complete the survey.

4. Discuss how bias can be introduced into statistics through:
 - a. questionnaire design
 - b. sample selection.

5. Explain how you can determine an appropriate sample size from a population of known size.

6. A number of people were asked to rate a movie on a scale of 0 to 5. Here are their scores:
 1, 0, 2, 1, 0, 0, 1, 0, 2, 3, 0, 0, 1, 0, 1, 2, 5, 3, 1, 0
 - a. Sort the data into a frequency distribution table.
 - b. Determine the mode.
 - c. Identify the median.
 - d. Calculate the range.

7. Weekly earnings from casual work performed by a sample of 50 high school students were rounded to the nearest dollar, as follows:



205 189 216 224 227 194 232 178 228 198 227 223 235 221 194 230 213
 226 241 220 179 235 186 208 194 208 223 238 226 234 219 219 197 225
 216 249 228 186 229 232 217 197 208 217 231 234 214 204 228 214

- a. Organise the data into a frequency distribution with class intervals $170- < 180$, $180- < 190$ and so on.
- b. Display the data as a histogram.
8. Calculate the mean of the following scores: 1, 2, 2, 2, 3, 3, 5, 4 and 6.
9. The mean of 10 scores was 5.5. Nine of the scores were 4, 5, 6, 8, 2, 3, 4, 6 and 9. Calculate the tenth score.
10. Consider the following distribution table.

Score (x)	Frequency (f)
2	3
3	2
5	8
6	2

Determine the:

- a. mean b. mode c. median d. range.
11. a. Determine the mode of the following values: 3, 2, 6, 5, 9, 8, 1, 7. Explain your answer.
 b. Determine the median of the following values: 10, 6, 1, 9, 8, 5, 17, 3.
 c. Calculate the range of the following values: 1, 6, 15, 7, 21, 8, 41, 7.

Problem solving

12. Consider the following distribution table.

Score (x)	Frequency (f)
1.5	10
2.0	20
2.5	8
3.0	5
3.5	6

- a. Calculate the mean score.
 b. Determine the median score.
 c. Give the modal score.
 d. Indicate which measure of centre best describes the distribution.
13. Consider the following stem plot.

Key: 6.1 | 8 = 6.18

Stem	Leaf
6.1	8 8 9
6.2	0 5 6 8
6.3	0 1 2 4 4 4

- a. Calculate the mean score.
 b. Determine the median score.
 c. Give the modal score.
 d. Indicate which measure of centre best describes the distribution.

14. Study this dot plot to answer the following questions.



- Calculate the mean score (correct to 2 decimal places).
 - Determine the median score.
 - Give the modal score.
 - Indicate which measure of centre best describes the distribution.
15. A frozen goods section manager recorded the following sales of chickens by size during a sample week.
16, 14, 13, 12, 15, 14, 13, 11, 12, 14, 14, 16, 15, 13, 11, 12, 14, 13, 15, 17, 13, 12, 14, 16, 13, 11, 15, 14, 12, 11, 15, 12, 13, 12, 12, 15, 13, 11, 11, 13, 16, 13, 12, 15, 17, 13, 14, 16, 12, 15
- Construct a frequency distribution table showing x , f and $f \times x$ columns. You may include a tally column if you wish.
 - Identify the mode of the distribution.
 - Calculate the mean and median sizes of the chickens sold.
 - Determine which size the manager should order most. Explain.
 - Calculate the range of sizes.
 - Calculate what percentage of total sales are in the 12–14 size group.
16. The following table displays the results of the number of pieces of mail delivered in a week to a number of homes.

Number of pieces of mail	Frequency
0	7
1	25
2	34
3	11
4	8
5	2
6	4
7	5
8	3
9	1

- Determine the most common number of pieces of mail delivered.
- Calculate the mean number of pieces of mail delivered.
- Calculate the range.
- Explain what this shows about the mail delivery service to these homes.



To test your understanding and knowledge of this topic, go to your learnON title at www.jacplus.com.au and complete the **post-test**.

Below is a full list of **rich resources** available online for this topic. These resources are designed to bring ideas to life, to promote deep and lasting learning and to support the different learning needs of each individual.

eWorkbook

Download the workbook for this topic, which includes worksheets, a code puzzle and a project (ewbk-1941)

Solutions

Download a copy of the fully worked solutions to every question in this topic (sol-0715)

Digital documents

- 10.4** SkillSHEET Presenting data in a frequency table (doc-6989)
- 10.5** SkillSHEET Finding the mean of ungrouped data (doc-6991)
- SkillSHEET Arranging a set of data in ascending order (doc-6992)
- SkillSHEET Finding the median (doc-6993)
- SkillSHEET Finding the middle score for data arranged in a dot plot (doc-6994)
- SkillSHEET Finding the score in a data set that occurs most frequently (doc-6995)

Video eLessons

- 10.2** Collecting data (eles-4434)
- Surveys (eles-4435)
- Planning a questionnaire (eles-4437)
- Selecting samples (eles-4438)
- Biased samples (eles-4439)
- 10.3** Primary data (eles-4440)
- Secondary data (eles-4441)
- 10.4** Examining data (eles-4442)
- Histograms (eles-4444)
- Using a spreadsheet to draw a histogram (eles-4445)
- 10.5** Mean (eles-4446)
- Median (eles-4467)
- Mode (eles-4468)
- Measures of spread (eles-4447)
- Clusters, gaps and outliers (eles-4448)
- Mean and median (eles-1905)
- 10.6** Analysing data sets (eles-4450)
- Using a sample to predict the properties of a population (eles-4451)

Interactivities

- 10.2** Individual pathway interactivity: Samples and populations (int-4451)
- Collecting data (int-3807)
- Questionnaires (int-3809)
- Planning a questionnaire (int-3810)
- Selecting samples (int-3811)
- Biased samples (int-3812)

- 10.3** Individual pathway interactivity: Primary and secondary data (int-4452)
- Primary and secondary data (int-3814)
- 10.4** Individual pathway interactivity: Organising and displaying data (int-4453)
- Frequency tables (int-3816)
- Column graphs (int-3817)
- 10.5** Individual pathway interactivity: Measures of centre and spread (int-4454)
- Mean (int-3818)
- Median (int-3819)
- Mode (int-3820)
- Range (int-3822)
- Outliers (int-3821)
- 10.6** Individual pathway interactivity: Analysing data (int-4456)
- 10.7** Crossword (int-2760)
- Sudoku puzzle (int-3191)

Teacher resources

There are many resources available exclusively for teachers online.

To access these online resources, log on to www.jacplus.com.au.

Answers

Topic 10 Representing and interpreting data

Exercise 10.1 Pre-test

- Primary data
- a. Mode = 3 b. Range = 10 c. Median = 4
- A and C
- B
- a. Mean = 0.86 b. Median = 1
- B
- 4
- Median
- B and D
- Mean = 1
- 165.5 cm
- A, B and C
- The mean would double.
- 3, 4, 5, 5, 5, 5
- $\frac{1}{4}$

Exercise 10.2 Samples and populations

- Census, sample
- Census — every member of the population participates
- Survey
- a. Survey
b. Survey
c. Census
d. Survey
- a. Survey
b. Census
c. Census
d. Survey
- Sample responses can be found in the worked solutions in the online resources.
- Sample responses can be found in the worked solutions in the online resources.
- a.-f. Sample responses can be found in the worked solutions in the online resources.
a.-c. Sample responses can be found in the worked solutions in the online resources.
- Example answers are shown here:
 - There is no correlation between choir members and using the school swimming pool. The sample is biased.
 - These people may or may not know about the PM's health-care package. In any case, the sample is biased as it is not truly representative of the population.
 - If the visitors had just arrived at the airport, they would not have had time to experience the transport system into the city. The sample is biased.
 - This sample size is far too small.

- Open questions have no boundaries for response. Closed questions require answers to fall within a category.
- One suggestion could be that the population would need to be 52 or fewer. Each person would be assigned a number and then a sample size of 10 would be selected from the deck of cards.
- 36, 40, 6, 16, 4, 25, 15, 43, 14, 50 is one possible solution.
- a. The population is all the students at the school.
b. Sample response: Every student in the school could be sent a survey to fill out about which social media platform they prefer.
c. Sample response: A sample can be selected by assigning each student a number and then randomly selecting a group of students using a random number generator.
- 6

Exercise 10.3 Primary and secondary data

- These are examples of simulations that could be conducted.
 - A coin could be flipped (Heads representing 'True' and Tails representing 'False').
 - A coin could be flipped (Heads representing 'red' and Tails representing 'black').
 - Spinner with 4 equal sectors (each sector representing a different toy)
 - Roll a die (each face represents a particular person).
 - Spinner with 3 equal sectors (each one representing a particular meal)
 - Spinner with 5 equal sectors (each one representing a particular destination)
- a. Some possible suggestions include:
Which students have internet access at home?
Do the students need access at night?
What hours would be suitable?
How many would use this facility?
b. Answers could include a survey or online questionnaire.
c. Sample responses can be found in the worked solutions in the online resources.
- Sample responses can be found in the worked solutions in the online resources.
- Some possible suggestions include:
Census, interview, observation, online response, experiment
- a. Measurement
b. Observation
c. Newspaper recordings
d. Survey
- The claim is false. It is not a logical deduction.
- Sample responses can be found in the worked solutions in the online resources.
- Pizza King's advertising campaign was misleading as it sounds as though all of their pizzas were rated 25% better than their competitors'. Only one of the 10 varieties was rated 25% better.
- Some possible suggestions include:
Primary data: Addison could visit each of the institutions in turn either in person or online.

Secondary data: She could seek advice from friends and colleagues. She could enlist the help of a mortgage broker or similar professional.

10. Sealy Posturepremier: 40% off $\left(\frac{1000}{2499} \times 100\%\right)$

Sealy Posturepedic: 41% off $\left(\frac{1600}{3899} \times 100\%\right)$

SleepMaker Casablanca: 40% off $\left(\frac{800}{1999} \times 100\%\right)$

SleepMaker Umbria: 42% off $\left(\frac{1800}{4299} \times 100\%\right)$

True; the discount is at least 40% off these beds.

11. a. Data set A

b. Data set B

c. Primary data is data that is collected by the researcher, and secondary data is data that has been collected by another source. Data set A was collected by Hannah, whereas data set B would have been found by another source.

12. A sample from one year group only will not give an indication of the preferences of the whole school. Hamish should ask students from all year levels.

13. Some possible suggestions include:

- The most-watched channel and the time slot with the highest viewing rating
- Conduct a survey in a supermarket to establish the typical customer for her product.
- Yes. If the typical customer was a five-year-old child, there would be little point in advertising during the 6 o'clock news, even though that time slot may have the highest rating on the most-watched channel.

Exercise 10.4 Organising and displaying data

1. a.

x	Frequency
0	5
1	8
2	13
3	7
4	5
5	2
Total	40

- The data are distributed fairly evenly around 2 children per household, and there appear to be no outliers. The graph clearly shows 2 children per household is the most common.
 - The sample is a random one, and of sufficient size; we can be confident that the suburb also exhibits these same properties.
2. a. The sample is a random one, so it seems to be a reliable reflection of the population of smart watches.

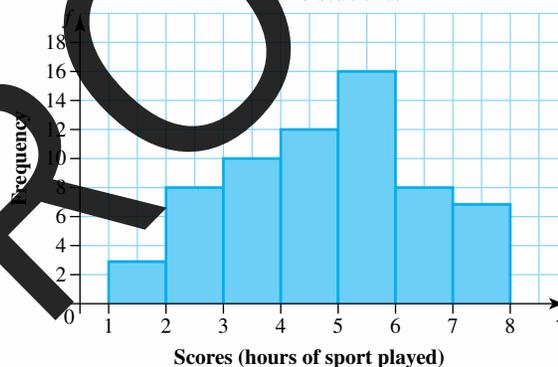
b.

x	Frequency
0	1
1	6
2	8
3	5
4	2
5	3
Total	25

- There was only 1 box with no defective smart watches. Most boxes had only 1, 2 or 3 defective smart watches, while 3 boxes were found to have 5 defective smart watches.
- Since the sample was randomly selected, it seems to be a reliable reflection of the characteristics of the population. It would be reasonably safe to say that most boxes would have only 1, 2 or 3 defective smart watches.

3. a.

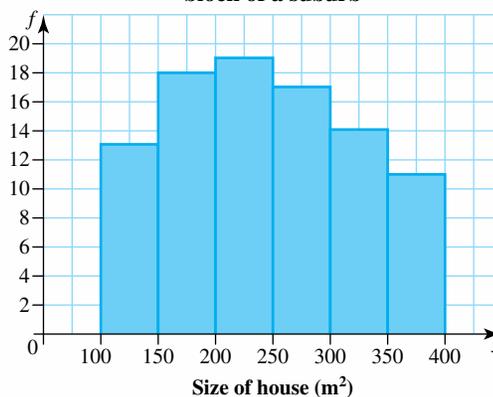
Hours of sport played by Year 8 students



- The graph rises steadily to a maximum, then falls away sharply at the upper end of the data.
- This is likely to be a true reflection of the sporting habits of Year 8 students. Some do a minimum of only 1 hour per week, quite a few do 2, 3 or 4 hours per week, with the maximum number doing 5 hours per week. The committed sports players would put in 6 or 7 hours per week.

4. a.

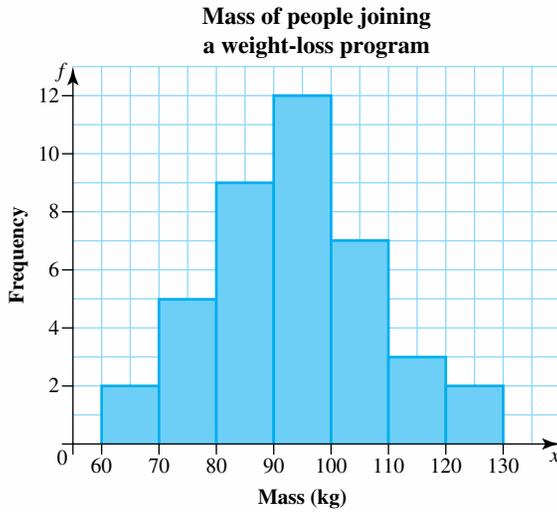
Size of houses within one block of a suburb



- The graph is roughly symmetrical, rising to a maximum at the 200-m² to 250-m² size, then decreasing slowly.

- c. Since this is one block of houses in the suburb, it is not a random sample. It is common for houses in a block of a suburb to be of similar style. For this reason, we could not say it reflects house sizes in the whole suburb.

5. a.

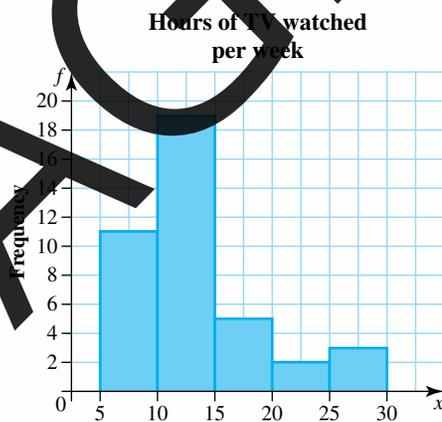


- b. The graph is quite symmetrical, rising to a maximum at the 90-kg to 100-kg mass, then decreasing more rapidly to the 130-kg mass.
- c. This would most likely not reflect the masses of people in the community because these are people who have enrolled in a program to lose weight.

6. a.

Class interval	Tally	Frequency
5- < 10		4
10- < 15		5
15- < 20		4
20- < 25		2
25- < 30		3
Total		18

b.



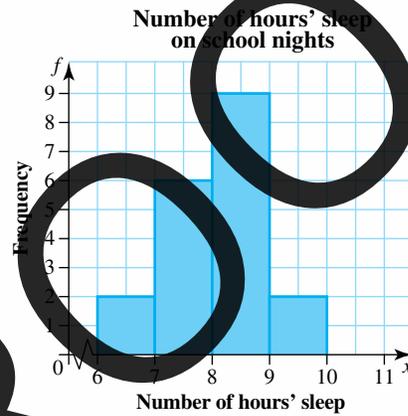
- c. The graph is heavily weighted towards the lower end of the scale, with most people watching fewer than 15 hours of TV per week. There were 3 people who watched almost 30 hours of TV per week.

- d. Since these people were interviewed in a shopping centre, the sample is not a random one. It could not therefore be taken to reflect the viewing habits of the community.

7. a.

Hours of sleep	Frequency
6- < 7	3
7- < 8	6
8- < 9	9
9- < 10	2
Total	20

b.

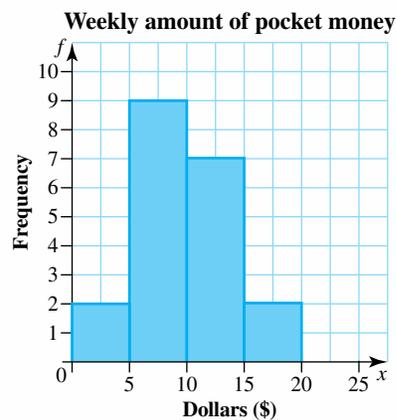


- c. The histogram peaks sharply at the 8 hours' sleep mark, indicating that generally, Year 8 students get 8 hours of sleep per night during the week. Some get less, and a few get more.
- d. It seems likely that these sample results would reflect the sleeping habits of Year 8 students generally.

8. a.

Pocket money (\$)	Frequency
0- < 5	2
5- < 10	9
10- < 15	7
15- < 20	2
Total	20

b.



- c. The histogram shows no general trend. The maximum is at \$10, indicating that a popular amount of pocket money is \$10 per week. Quite a few receive less than this, with only 2 receiving more.

- d. Since this is a random sample, it is quite likely that these results reflect the general population of 13-year-olds when it comes to pocket money.

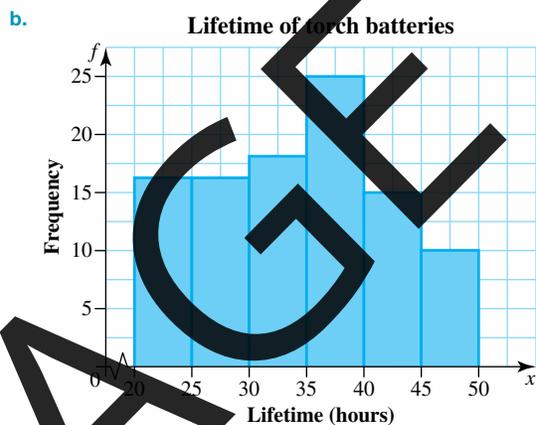
9. $\frac{12.5 + 13.2}{2} = 12.85$

10.

Class interval	Frequency
20–<30	2
30–<40	4
40–<50	4
50–<60	5
60–<70	3
70–<80	1
80–<90	0
90–<100	1

11. a.

Lifetime (hours)	Frequency
20–<25	16
25–<30	16
30–<35	18
35–<40	25
40–<45	15
45–<50	10
Total	100



- c. The histogram shows that the majority of torch batteries last for about 40 hours. A few last longer than this.
 d. It seems reasonable that the torch battery population would display a similar trend.

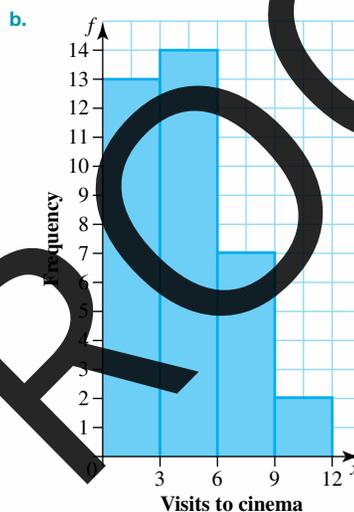
12. a. 56

b. 9–12

- c. No, it is not possible to determine the maximum number of weekends worked. However, the highest interval is 25–28. Therefore, the most weekends worked may have been 28; but we cannot be certain whether any tradesperson actually worked 28 weekends.

13. a. Answers will depend on class intervals chosen. An example is given.

Number of visits to the cinema	Tally	Frequency (f)
0–2		13
3–5		14
6–8		7
9–11		2
Total		36



c. 13

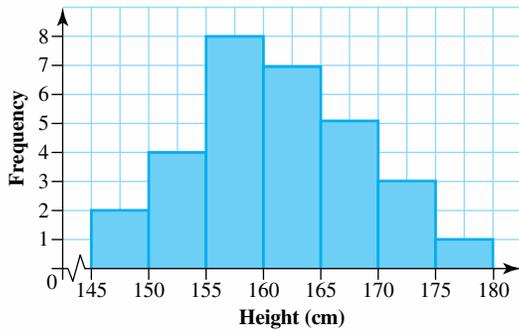
d. 23

- e. No. There is no information to explain how the sample of people from a particular population was obtained. Before any conclusions can be drawn, we must know what the population was and if the sample was random.

14. a.

Class interval	Tally	Frequency
145–<150		2
150–<155		4
155–<160		8
160–<165		7
165–<170		5
170–<175		3
175–<180		1
Total		30

b. Heights of Year 9 students



c.

Class interval	Tally	Frequency
145- < 148		1
148- < 151		2
151- < 154		2
154- < 157		4
157- < 160	/	5
160- < 163		4
163- < 166		3
166- < 169	/	5
169- < 172		2
172- < 175		1
175- < 178	—	0
178- < 181		1
Total		30

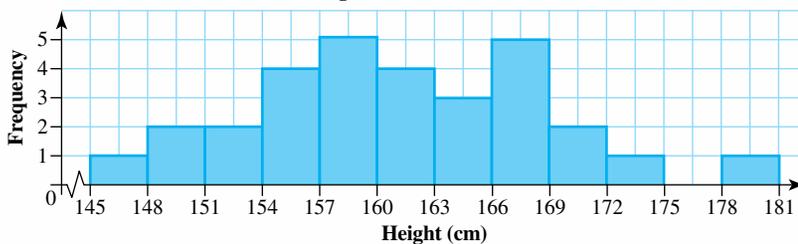
- d. See the graph at the foot of the page.*
- e. The two histograms represent the same data set, but appear to be quite different. The first histogram appears to be roughly symmetrical, with a maximum number of students having a height of about 155–160 cm. The second histogram has two modes, with the most common height for students being about 157 cm to 160 cm, or 166 cm to 169 cm. It illustrates the fact that the interpretation of a histogram displaying grouped data is dependent on the class interval used.

Exercise 10.5 Measures of centre and spread

- 26.8
 a. 4.33 b. 7.33 c. 36.30 d. 6.91

*14. d.

Heights of Year 9 students



3. a. 18 b. 10 c. 1.8 goals
 4. C
 5. Group A: 6.1; Group B: 6.8. Group B has a higher mean.
 6. 2.29
 7. 7.90
 8. a. 7 b. 17
 9. a. 50 b. 1.75
 10. Question 8:
 a. 5
 b. 17
 Question 9:
 c. 52
 d. 1.5, 1.7, 1.8

11. C
 12. E
 13. a. 8 b. 15 c. 50 d. 2.5
 14. a. 4 b. 5 c. 15 d. 4
 15. a. 35 b. 46 c. 35
 16. Mean = 6; median = 6; mode = 6; range = 4
 17. a. 6.7 b. 10

18.

x	f	f × x
0	1	0
1	2	2
2	7	14
3	3	9
4	4	16
5	2	10
6	1	6
Total	20	57

Mean = 2.85

19. a. 36 b. 7
 20. a. 3 b. 3 c. Both
 21. a. 49 b. 55
 c. 56 d. Median
 22. a. \$18 000
 b. \$26 000
 c. \$30 000
 d. Mean (the highest value)
 e. Mode (the lowest value)
 23. a. 8 b. No
 24. 31–40
 25. B

26. a. Mean = 26.6; median = 27.5; mode = 30; range = 22
 b. Mean = 32; median = 29; mode = 30; range = 70
 c. The inclusion of an outlier dramatically increased the range and significantly increased the mean of the data. The median was marginally increased while the mode was unchanged.
27. a. Mean = 4.5; median = 5; modes = 3 and 5
 b. The median is higher than the mean.
 c. The mean is between the two modal values.
 d. As there are no outliers, the mean is probably the best indicator of the centre of the data set.
28. a. The median, as there are likely to be some outliers which will significantly affect the mean
 b. The mode, as this will be the most popular shoe size and will be a number that is easy to interpret
 c. The mean, as this will be the average scored in a game
29. a. The potential error is the data value of 23.
 b. This value could have been a genuine outlier if the particular student was a multi-sport athlete who trained for a number of hours each day.
30. 73.45
31. Mean = 5; median = 3; mode = 6
32. Highest score is 49 and lowest score is 0.
 The scores would be 0, 1, 12, 13, 49.
33. 171

Exercise 10.6 Analysing data

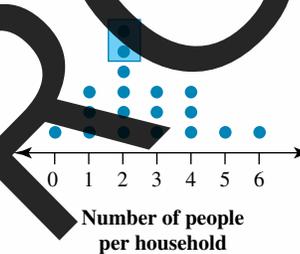
1. a. Range
 c. Mean
 2. a. 6.4
 b. 6
 c. 6
 d. They are all quite close, so any would do.
 e. Since these were the results of half the class, and the sample was random, it seems likely that the population results would be similar.
3. a. i. 2
 iii. 1
 b. i. 7.8
 iii. 7
 4. a. i. 22.6
 iii. 23
 b. i. 117.5
 iii. 112
5. a. i. 4.5
 ii. 5
 iii. 5
 iv. Any of the three
 b. i. 82.4
 ii. 82
 iii. 81
 iv. Mean or median
- b. Mode
 d. Median
- ii. 2
 v. Mean or median
- ii. 7
 iv. Any of the three
- ii. 23
 iv. Any of the three
- ii. 116
 iv. Mean

6. a. Sample 1: mean = 107.9; median = 101.5; no mode; range = 137
 Sample 2: mean = 112.5; median = 106.5; mode = 85; range = 127
 Sample 3: mean = 125.4; median = 104; no mode; range = 190
 b. All three median values lie close together. Sample 3 has a significantly larger mean and range than the other two samples, possibly caused by the data value 236, which appears to be an outlier. Sample 2 is the only sample that has a mode, so modes cannot be compared.

7. a.

Score (x)	Frequency (f)
0	1
1	3
2	6
3	3
4	3
5	1
6	1
	$n = 18$

$\bar{x} = 2.6$



Median = 2

- c. 2
 d. i. Median ii. Mean iii. Mode

8. a.

Score (x)	Frequency (f)
137	2
138	3
139	5
140	3
141	6
142	1
	$n = 20$

- b. 141, 139.5, 139.6
 c. Mean
9. a. The median was calculated by taking the average of the 2 middle scores.
 b. 13
10. 2
11. a. 55 250 km, 52 000 km, 52 000 km
 b. The mean, as the data did not appear to have any outliers
 c. 51 810 km; it is reduced by 3440 km.
 d. 52 000 km
 e. 3333
 f. 50 000 km; 92% last that distance or more.

10. a. 4.3 b. 5 c. 5 d. 4
11. a. There is no mode since none of the values occurs more than once.
 b. 7
 c. 40
12. a. 2.3
 b. 2.0
 c. 2.0
 d. Mode
13. a. 6.27
 b. 6.28
 c. 6.34
 d. Mean or median
14. a. 2.63
 b. 2.55
 c. 2.4
 d. Mean

15. a.

x	f	$f \times x$
11	6	66
12	10	120
13	11	143
14	8	112
15	8	120
16	5	80
17	2	34
Total	50	675

- b. 13
 c. 13.5, 13
 d. 13; as this is the most frequently sold size
 e. 6
 f. 58%
16. a. 2
 b. 2.6
 c. 9
 d. Most homes get up to about 3 pieces of mail. Some do get more.