Mathematics Level 7

Chapter 6 — Statistics

Attribution: Classbasics Pty Ltd, vwx.au

Aboriginal and Torres Strait Islander content has been excluded as accuracy cannot be guaranteed, and it is better to be respectful than cover the entire curriculum inaccurately.

Significant effort has been made to ensure this content is legal. Please send any areas of concern to admin@vwx.au with appropriate document/legal references so errors can be verified and changes be made.

Contents

Section	Page
Acquiring data; range, median, mean and mode	2
Dot plots and stem-and-leaf plots: shape, centre and spread	13
Planning and conducting a statistical investigation	27
Test	38

Acquiring data; range, median, mean and mode

Content description `VC2M7ST01`: acquire data sets for discrete and continuous numerical variables and calculate the range, median, mean and mode; make and justify decisions about which measures of central tendency provide useful insights into the nature of the distribution of data

Achievement standard (extract): Students plan and conduct statistical investigations involving discrete and continuous numerical data, using appropriate displays.

Theory

Data and variables. Some questions are answered by collecting data — and the data will *vary* from one person or thing to the next. The thing being asked about is the **variable**, and the collection of values is the **data set**. This subtopic works with **numerical** data: values that are numbers. Subtopic 6C turns this into a full investigation; here we learn to get a data set and then summarise it with four useful numbers.

Two kinds of numerical variable. A **discrete** variable is counted: its values are whole numbers (number of pets, goals kicked in a match). A **continuous** variable is measured: it can take any value within a range, and you can only write it down to the smallest mark on your ruler, stopwatch or scale (height, time, temperature). The kind of variable decides how you acquire the data and what the values can look like — see Figure 1.

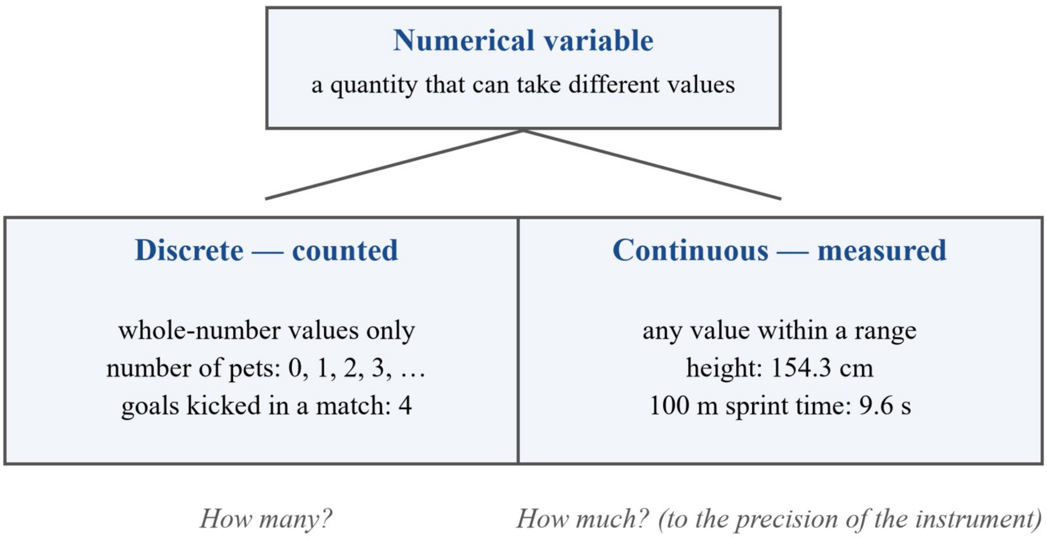

Figure 1. A tree diagram: a numerical variable — a quantity that can take different values — splits into discrete (counted, whole-number values only: number of pets 0, 1, 2, 3, …; goals kicked in a match 4) and continuous (measured, any value within a range: height 154.3 cm; 100 m sprint time 9.6 s).

Acquiring data. There are three ways to get a data set, matching the kind of variable (Figure 2). You **count** discrete values — a tally, a survey of the class. You **measure** continuous values — and a science experiment, observation or field study is one of the most common ways: the time a marble takes to roll down a ramp, the temperature of cooling water, the length of leaves from the same plant. Or you **find** data that someone else has already collected and published — the Australian Bureau of Statistics and the Bureau of Meteorology publish data sets far bigger than a class could collect. (Which source to choose, and how to plan the collection, is 6C's job; here the point is that every data set arrives by one of these three doors.)

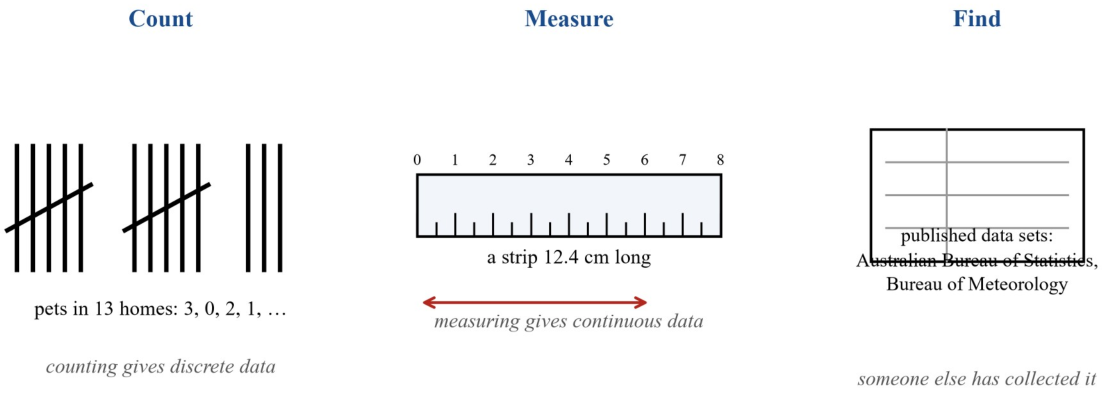

Figure 2. Three panels showing how a data set is acquired. Count: a tally of 13, giving discrete data. Measure: a ruler measuring a strip 12.4 cm long, giving continuous data. Find: a published table from the Australian Bureau of Statistics or the Bureau of Meteorology, collected by someone else.

Putting the data in order. Nearly everything in this subtopic gets easier once the values are written in order, smallest to largest. For the nine values 12, 3, 7, 5, 3, 9, 5, 7, 7 the ordered list is 3, 3, 5, 5, 7, 7, 7, 9, 12, and the middle and the two ends are suddenly easy to see.

The range — how spread out? The **range** is the simplest number about spread:

range = largest value − smallest value.

For 3, 3, 5, 5, 7, 7, 7, 9, 12 the range is $12 - 3 = 9$. The range uses only the two extreme values, so one unusual value can make it large even when most of the data sit close together.

The median — the middle. Order the values, then the **median** is the middle value. If there is an odd number of values, one value sits in the exact middle. If there is an even number, two values share the middle, and the median is their mean. Crossing values off in pairs from the two ends finds the middle reliably (Figure 3).

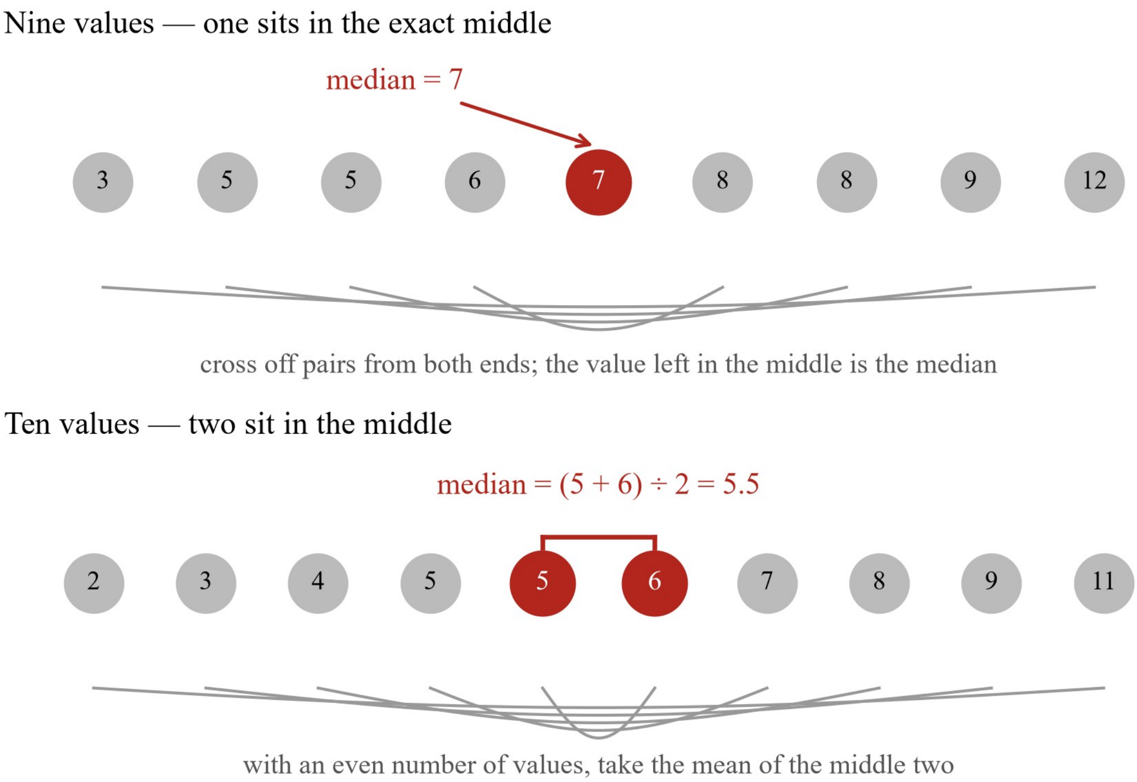

Figure 3. Two rows of dots with the values written in them. Top: nine ordered values 3, 5, 5, 6, 7, 8, 8, 9, 12 with pairs crossed off from both ends, leaving the middle value 7 as the median. Bottom: ten ordered values 2, 3, 4, 5, 5, 6, 7, 8, 9, 11 with the middle two, 5 and 6, marked, giving a median of (5 + 6) ÷ 2 = 5.5.

The mean — sharing it out evenly. Add up every value, then divide by how many values there are:

mean = sum of the values ÷ number of values.

For the books read by five students — 3, 5, 4, 7, 6 — the sum is 25 and the mean is $25 \div 5 = 5$. Picture the values as bars: the mean is the height the bars would stand at if the total were levelled out evenly (Figure 4). Because the mean adds up *every* value, one very big or very small value drags it.

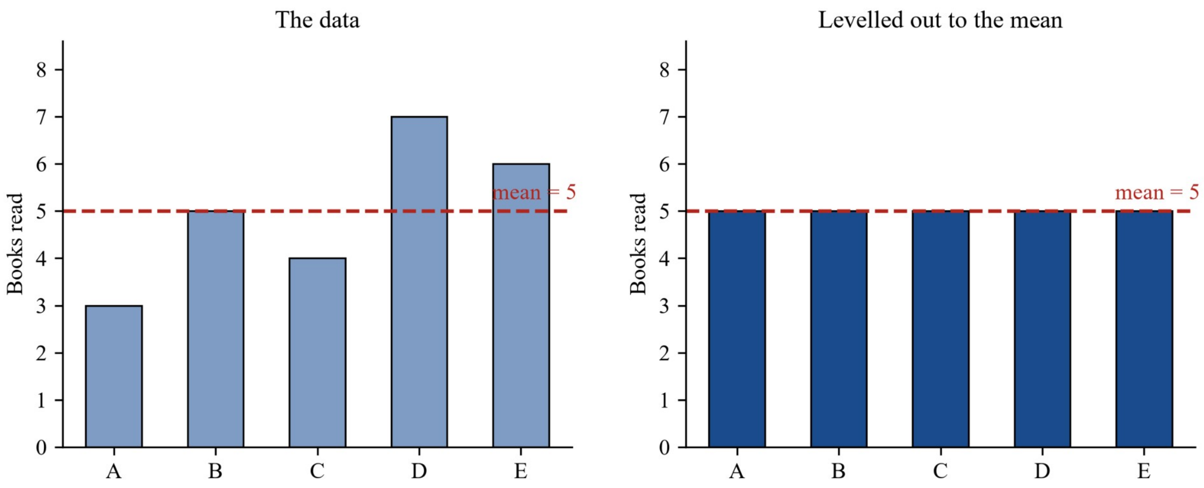

Figure 4. Two bar charts of the books read by five students (3, 5, 4, 7, 6). Left: the bars at their own heights with a dashed line at 5. Right: the same total levelled out so every bar is 5 high — the mean.

The mode — the most common. The **mode** is the value that appears most often. It needs no ordering, only counting: in 3, 3, 5, 5, 7, 7, 7, 9, 12 the mode is 7. If no value repeats, the data set has no mode at all.

Four numbers, two jobs. The range says how spread out the data are; the median, mean and mode each give a different idea of the centre — a "typical" value. Together they are called **measures of central tendency**, and choosing which one is useful in a given situation is the real skill in this subtopic.

Outliers. The Victorian Curriculum uses the word *outlier* but does not define it, so this book states its own working definition: an **outlier** is a value that sits a long way from the rest of the data — far enough away that you would check it before trusting it. In Figure 5 the 24 is an outlier among the values 5 to 9. With it, the mean is $87 \div 10 = 8.7$; without it, the mean of the remaining nine values is $63 \div 9 = 7$. The median is 7 either way. That is the whole story in one picture: **the mean is pulled towards an outlier; the median does not move.**

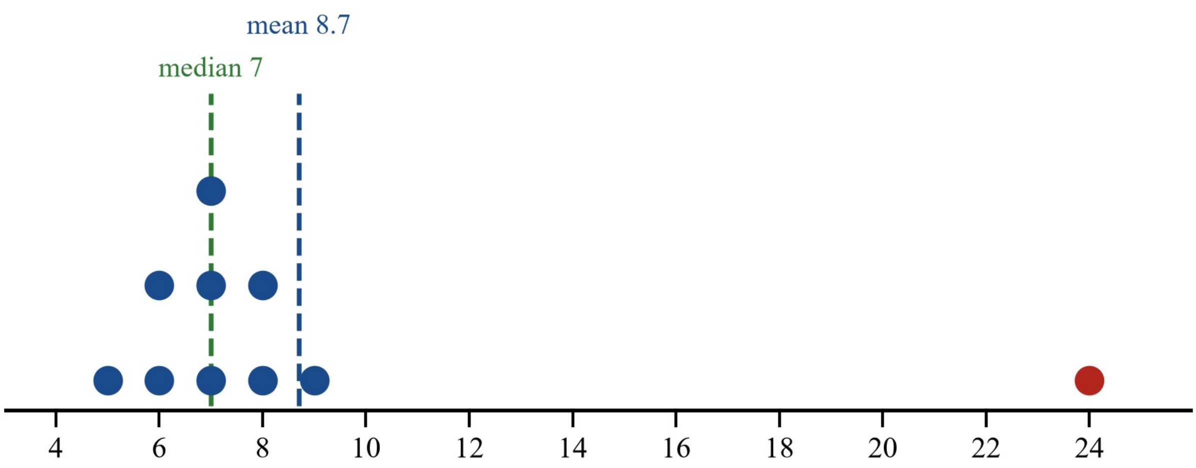

Figure 5. Dot plot of the ten values 5, 6, 6, 7, 7, 7, 8, 8, 9, 24 with the 24 drawn in red as an outlier; a dashed line marks the median at 7 and a second dashed line marks the mean at 8.7, pulled towards the outlier.

Which centre is useful? Different data deserve different centres, and saying *why* is the decision the curriculum asks for. As a guide:

- **No outlier, and the values sit evenly:** the mean and median are close, and the mean is a good choice because it uses every value.
- **An outlier, or a long tail on one side:** the median is usually the truer "typical" value, because the mean is dragged towards the tail. (6B gives these shapes names, and draws them.)
- **The question is about what is most common** — which size to make more of, which choice is most popular — the mode answers it directly, and the mean and median can both miss the point.

Figure 6 shows the decision on real data. The daily rainfall at Melbourne (Olympic Park) for 1–18 August 2026 is nearly all zero — eleven dry days — with a few wet days and one storm day of 24.4 mm. The mean is $41.4 \div 18 = 2.3$ mm, dragged up by the storm day; the median is 0 mm; the mode is 0 mm. "A typical mid-August day in Melbourne is dry" is what the median and mode say, and it is true; the mean of 2.3 mm describes no day at all. Here the median is the useful centre, and the range of 24.4 mm warns how much the rain varied from day to day.

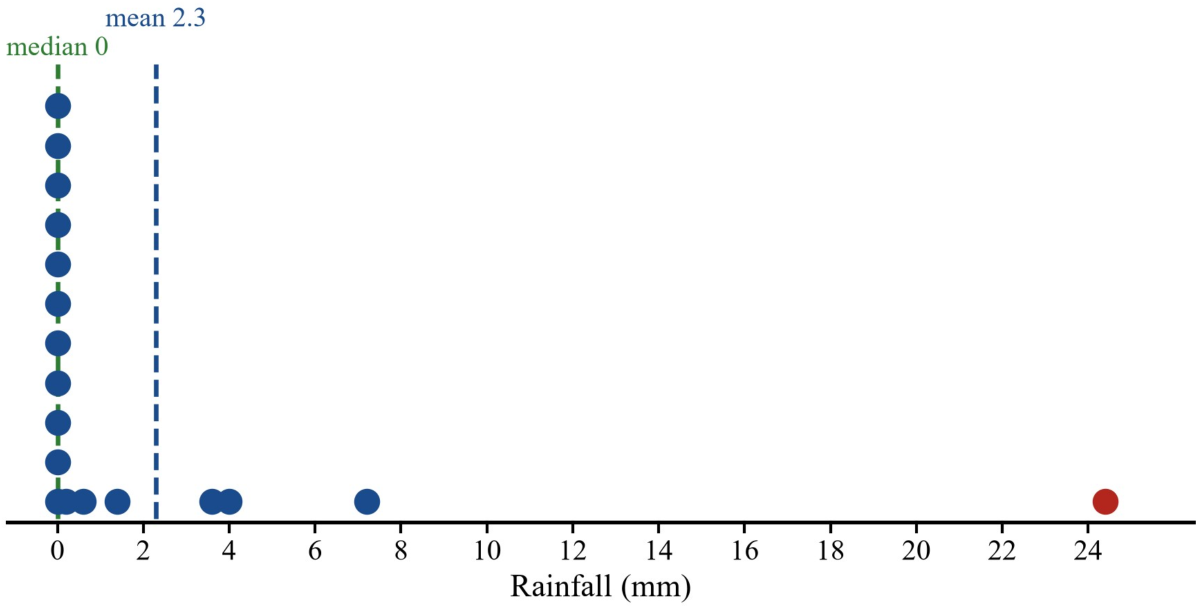

Figure 6. Dot plot of the daily rainfall at Melbourne (Olympic Park) for 1–18 August 2026: eleven dots stacked at 0, single dots at 0.2, 0.6, 1.4, 3.6, 4.0 and 7.2, and the storm day of 24.4 drawn in red. Dashed lines mark the median at 0 and the mean at 2.3.

Source: Bureau of Meteorology, "Daily Weather Observations, Melbourne (Olympic Park), Vic" (station 086338), 1–18 August 2026, IDCJDW3033.202608, prepared 19 August 2026; retrieved 20 August 2026.

Same centre, different data. Two data sets can share their summary numbers and still be different sets. Sets A (6, 7, 8, 8, 8, 9, 10) and B (2, 4, 8, 8, 8, 12, 14) in Figure 7 both have mean 8, median 8 and mode 8 — yet A clusters close to 8 while B spreads out to the ends (range 4 against range 12). Changing the outer values changes the range and does not move the centre; changing the middle values moves the centre and leaves the range alone. Summary numbers are a summary, not a photograph: the centre and the spread are reported *together* because neither tells the whole story.

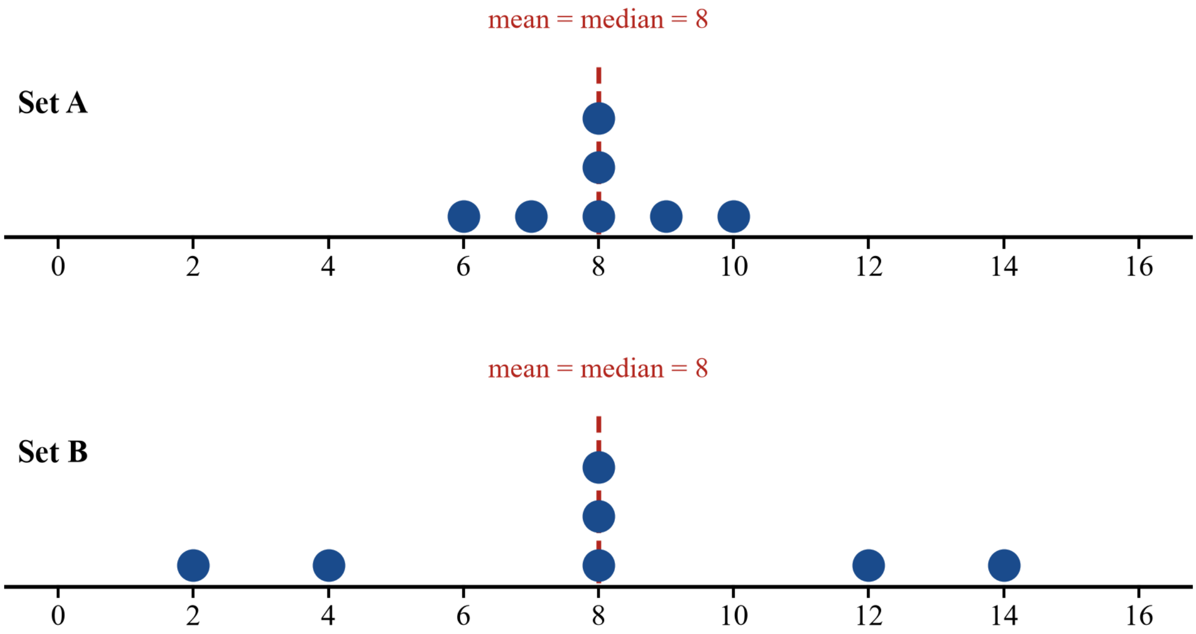

Figure 7. Two dot plots on the same number line. Set A: 6, 7, 8, 8, 8, 9, 10. Set B: 2, 4, 8, 8, 8, 12, 14. Both have a dashed line where mean and median coincide at 8, but Set A clusters close to 8 while Set B spreads out to the ends.

The words of this subtopic, with synonyms you may meet:

Term	Meaning here	Synonyms and related words
variable	the thing being measured or counted, whose values vary	—
data set	the collection of values of a variable	data
discrete	counted; whole-number values	—
continuous	measured; any value within a range	—
range	largest value − smallest value	spread
median	the middle value once ordered; with an even number of values, the mean of the middle two	middle value
mean	sum of the values ÷ number of values	average
mode	the value that appears most often	most common value
measures of central tendency	the mean, median and mode together	measures of centre
outlier	a value a long way from the rest of the data	anomaly

Worked examples

Example 1 — scaffolded

The daily maximum temperatures (°C) at Melbourne (Olympic Park) for the first 18 days of August 2026 were:

16.6, 17.6, 15.9, 13.6, 13.4, 16.8, 16.0, 17.7, 14.3, 13.1, 13.3, 15.6, 15.0, 16.7, 14.1, 14.4, 14.1, 20.0

Source: Bureau of Meteorology, "Daily Weather Observations, Melbourne (Olympic Park), Vic" (station 086338), 1–18 August 2026, IDCJDW3033.202608, prepared 19 August 2026; retrieved 20 August 2026.

Find the range, median, mean and mode, then say which centre describes a typical day.

Working	Comment
Order the values: 13.1, 13.3, 13.4, 13.6, 14.1, 14.1, 14.3,	Ordering first makes every later step easier.

Working	Comment
14.4, 15.0, 15.6, 15.9, 16.0, 16.6, 16.7, 16.8, 17.6, 17.7, 20.0.	
Range $= 20.0 - 13.1 = 6.9$ °C.	Largest minus smallest.
Median: $n = 18$, so average the 9th and 10th values: $(15.0 + 15.6) \div 2 = 15.3$ °C.	An even number of values puts the median halfway between the middle two.
Mean: the sum is 278.2, and $278.2 \div 18 \approx 15.5$ °C.	The mean uses every value.
Mode $= 14.1$ °C, the only value that repeats (twice).	Counting repeats finds the mode.
The mean (15.5 °C) and the median (15.3 °C) are close: the values sit evenly, with only a small tail at 20.0 °C.	When the mean and median nearly agree, either is a fair "typical" value.

∴ a typical day reached about 15–16 °C; with no outlier, the mean (15.5 °C) and the median (15.3 °C) tell the same story.

Example 2 — scaffolded

The daily rainfall (mm) at the same station over the same 18 days was:

0, 0, 0, 0.6, 3.6, 7.2, 0.2, 0, 1.4, 24.4, 4.0, 0, 0, 0, 0, 0, 0, 0

Find the range, median, mean and mode, identify any outlier, and decide which centre is useful here. (The data are drawn in Figure 6.)

Working	Comment
Order the values: eleven 0s, then 0.2, 0.6, 1.4, 3.6, 4.0, 7.2, 24.4.	The zeros are data too — a dry day is a measurement of 0 mm.
Range $= 24.4 - 0 = 24.4$ mm.	One storm day sets the range.
Median: $n = 18$; the 9th and 10th values are both 0, so the median is 0 mm.	The middle of the ordered list is dry.
Mean: the sum is 41.4, and $41.4 \div 18 = 2.3$ mm.	Every value is added, including the storm.
Mode $= 0$ mm (eleven days).	The most common day is a dry day.
Outlier: 24.4 mm — the next-largest value is only 7.2 mm. Without it, the mean of the other 17 days is $17.0 \div 17 = 1.0$ mm; the median is still 0 mm.	The outlier dragged the mean up by 1.3 mm and did not move the median.

∴ the median (0 mm) and the mode (0 mm) are the useful centres: a typical mid-August day was dry. The mean of 2.3 mm is pulled up by the one storm day and describes no actual day.

Example 3 — unscaffolded

In a science investigation, a Year 7 class timed how long (s) a marble took to roll down the same ramp, releasing it the same way each trial:

3.2, 3.4, 3.1, 3.3, 3.2, 3.5, 3.3, 3.2, 3.4, 3.3, 5.8

The dot plot is shown in Figure 8. Identify any anomaly in the data, find the median and the mean with and without it, and say which centre the class should report.

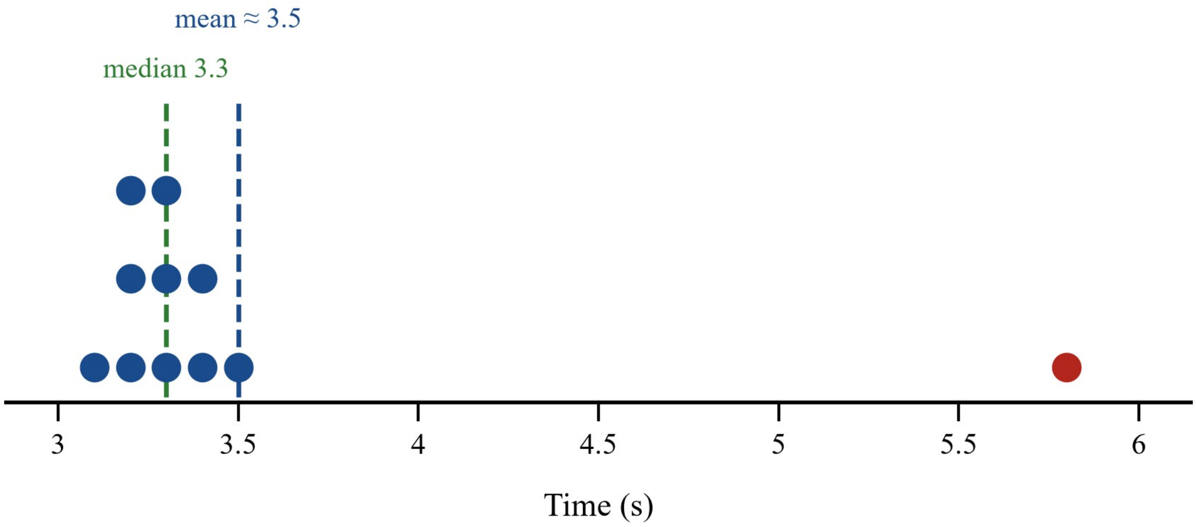

Figure 8. Dot plot of eleven marble roll times from 3.1 s to 3.5 s with the anomaly of 5.8 s drawn in red; dashed lines mark the median at 3.3 and the mean at about 3.5.

Ordered, the ten close values run from 3.1 s to 3.5 s, and the 5.8 s trial sits a long way from them — it is an outlier (perhaps the marble caught on a join in the ramp). With all eleven values, the median (6th value) is 3.3 s and the mean is $38.7 \div 11 \approx 3.5$ s. Without the 5.8 s trial, the mean of the ten remaining values is $32.9 \div 10 \approx 3.3$ s and the median is still 3.3 s. The one odd trial dragged the mean up to the edge of the data and left the median in the middle of the pack.

∴ the class should report the median, 3.3 s, as the typical roll time: the mean with the outlier (3.5 s) is slower than ten of the eleven trials, while the median is what the marble actually did.

Example 4 — unscaffolded

Two players' goals across seven matches were:

Player X: 6, 9, 11, 12, 13, 15, 18 Player Y: 11, 11, 12, 12, 12, 13, 13

Show that both players have the same mean and the same median, then explain how the two data sets still differ, and which player a coach would pick for a steady score.

Both sums are 84, so both means are $84 \div 7 = 12$ goals; both middle (4th) values are 12, so both medians are 12 goals. The spread differs: X's range is $18 - 6 = 12$ goals, Y's is $13 - 11 = 2$ goals. X swings between quiet matches and big ones; Y is always near 12.

∴ for a steady score the coach picks Y — the same centre, but a far smaller range. Mean and median alone cannot tell the players apart; the range can.

Practice questions

Scaffolded

Q1. Say whether each variable is discrete or continuous. (a) The number of pets in a student's home. (b) The height of a student (cm). (c) The time taken to walk to school (minutes). (d) The number of apps on a student's phone. (e) The mass of a school bag (kg). (f) The number of goals a team kicks in a match.

Q2. Nine hens laid these numbers of eggs in one week:

3, 4, 4, 5, 5, 5, 6, 6, 7

(a) Find the mode.
(b) Find the median.
(c) Find the range.
(d) Find the mean.

Q3. Ten students scored these marks out of 10 on a spelling test:

5, 6, 6, 7, 7, 7, 8, 8, 9, 10

(a) Find the mode.
(b) Find the median.
(c) Find the range.
(d) Find the mean.

Q4. A shop records the size of every pair of soccer boots it sells in a month, so that it knows which size to stock more of. Which measure should it look at?

A. the mean B. the median C. the mode D. the range

Q5. Seven parcels on a desk have masses (kg):

4, 5, 6, 6, 7, 8, 40

(a) Find the median and the mean.
(b) Which value is an outlier?
(c) Find the mean again without the outlier. Which centre — the mean or the median — did the outlier change more?

Q6. For each data set below, say (i) whether the variable is discrete or continuous, and (ii) whether you would acquire the data by counting, by measuring, or by finding a published source. (a) The number of eggs each hen in a pen lays in a week. (b) The mass of each apple in a tray (g). (c) The maximum temperature in Melbourne yesterday (°C).

Unscaffolded

Q7. The daily minimum temperatures (°C) at Melbourne (Olympic Park) for the first 18 days of August 2026 were:

2.6, 5.0, 8.2, 7.6, 9.4, 8.1, 10.3, 9.8, 11.5, 10.1, 9.7, 9.5, 9.0, 11.6, 8.9, 6.1, 7.3, 6.8

Source: Bureau of Meteorology, "Daily Weather Observations, Melbourne (Olympic Park), Vic" (station 086338), 1–18 August 2026, IDCJDW3033.202608, prepared 19 August 2026; retrieved 20 August 2026.

Find the range, median and mean (to 1 decimal place). Explain why this data set has no mode, and say whether the mean and median tell the same story here.

Q8. Two data sets are:

A: 3, 8, 8, 8, 13 B: 3, 5, 8, 11, 13

Find the mean, median and range of each set. What do the two sets share, and how do they differ? What does that say about summarising data with numbers alone?

Q9. The prices of seven houses sold in a suburb this year were (in thousands of dollars):

500, 520, 540, 560, 580, 600, 2300

(a) Find the mean and the median.
(b) An advertisement says "average price $800 000". Which measure is that?
(c) Which measure gives a buyer the truer picture of a typical house price in the suburb? Why?

Q10. Figure 9 shows the number of people living in the homes of 15 students.

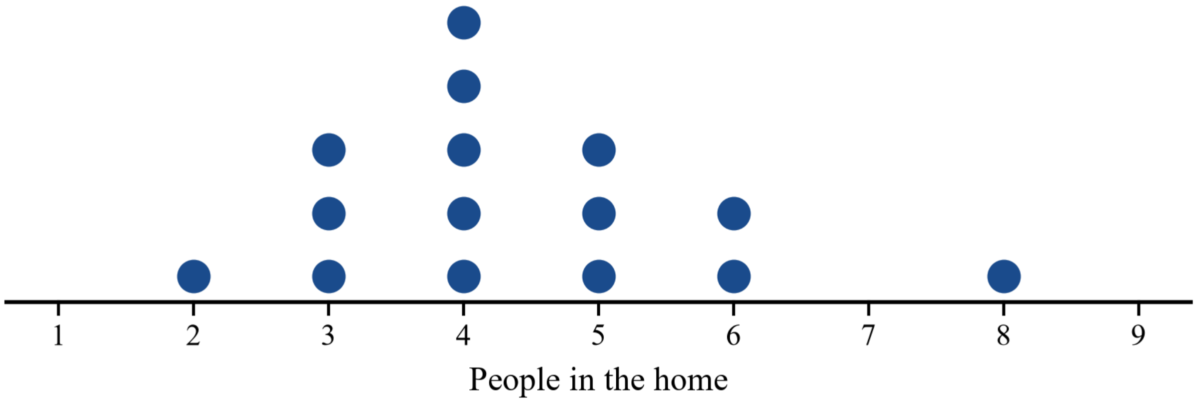

Figure 9. Dot plot of the number of people in the homes of 15 students: one dot at 2, three at 3, five at 4, three at 5, two at 6 and one at 8.

(a) Read the fifteen values off the plot and find the mode, the median, the range and the mean.
(b) Which centre best describes a typical home in this survey? Why?

Q11. Five values have a mean of 10. Four of them are 8, 9, 12 and 13. (a) Find the fifth value. (b) With all five values in order, find the median and the mode.

Q12. A canteen recorded the number of sandwiches sold on each of ten school days:

40, 42, 44, 45, 46, 47, 48, 50, 52, 118

(The 118 was the day of the whole-school sports carnival.) Find the mean and the median, and say which number the canteen should use to plan how many sandwiches to make on an ordinary day.

Q13. Write down a data set of five values with median 6, mode 6 and range 8. Check your answer by calculating the three numbers. (There is more than one correct answer.)

Q14. A class wonders whether Year 7 students jump further than Year 5 students in a standing long jump. (a) Name the variable the class would measure, and its units. (b) Is the variable discrete or continuous? (c) Describe how the class could acquire the data. (d) Name one centre the class could compare between the two year levels, and say what an outlier in one group would do to it.

Q15. The ages of the nine people at a junior training session were:

7, 8, 9, 9, 9, 10, 10, 11, 34

(The 34 is the coach.) Find the mean (to 1 decimal place), the median, the mode and the range, and say which centre best describes a typical person at the session.

Q16. Five values have a mean of 8. Four of them are 5, 7, 9 and 10. (a) Find the fifth value. (b) The 10 is now replaced by 20. Without adding all the values again, find the new mean. (c) Find the median before and after the change, and explain in one sentence why the mean moved but the median did not.

Answers

Q1. (a) discrete. (b) continuous. (c) continuous. (d) discrete. (e) continuous. (f) discrete.

Q2. (a) 5 eggs. (b) 5 eggs. (c) 4 eggs. (d) 5 eggs.

Q3. (a) 7 marks. (b) 7 marks. (c) 5 marks. (d) 7.3 marks.

Q4. C — the mode.

Q5. (a) Median 6 kg; mean ≈ 10.9 kg. (b) 40 kg. (c) Mean without it: 6 kg. The outlier changed the mean (from 10.9 down to 6) and did not change the median.

Q6. (a) (i) discrete; (ii) counting. (b) (i) continuous; (ii) measuring. (c) (i) continuous; (ii) finding a published source (Bureau of Meteorology).

Q7. Range 9.0 °C; median 8.95 °C; mean ≈ 8.4 °C; no mode because no value repeats. The mean and median tell much the same story (both about 8–9 °C), with the one cold night (2.6 °C) pulling the mean a little below the median.

Q8. A: mean 8, median 8, range 10. B: mean 8, median 8, range 10. They share all three numbers yet are different sets — A bunches at 8, B spreads evenly. Summary numbers alone do not pin down the data.

Q9. (a) Mean 800 (thousand dollars); median 560 (thousand dollars). (b) The mean. (c) The median — six of the seven houses sit between 500 and 600 thousand, and the one very expensive house drags the mean far above them.

Q10. (a) Mode 4 people; median 4 people; range 6 people; mean 4.4 people. (b) The median (or the mode), 4 — the one home with 8 people pulls the mean above what most homes hold.

Q11. (a) 8. (b) Median 9; mode 8.

Q12. Mean 53.2; median 46.5. Use the median — about 46–47 sandwiches — because the carnival day (118) drags the mean above every ordinary day.

Q13. Example: 2, 4, 6, 6, 10 — median 6, mode 6, range $10-2=8$.

Q14. Example answers. (a) The length of each student's best standing long jump, in cm. (b) Continuous — it is measured. (c) Measure each student's best of two or three jumps, in both year levels, with the same method. (d) The median — and an outlier in one group would drag that group's mean towards it but would not move its median.

Q15. Mean ≈ 11.9; median 9; mode 9; range 27. The median (or mode), 9 — the coach's age drags the mean above every student's age.

Q16. (a) 9. (b) The sum grows by 10, so the new mean is $50\div5=10$. (c) Median 9 before and after — the middle value of the ordered set does not change when the largest value grows; the mean moved because the sum grew.

Fully-worked solutions

Q1. Counted quantities take whole numbers — (a), (d) and (f) are discrete. Measured quantities take any value within a range — (b), (c) and (e) are continuous.

Q2. The values are already ordered. (a) 5 appears three times, more than any other value: mode 5 eggs. (b) With nine values the median is the 5th: 5 eggs. (c) Range $=7-3=4$ eggs. (d) The sum is 45, so the mean is $45\div9=5$ eggs.

Q3. (a) 7 appears three times, more than any other value: mode 7 marks. (b) With ten values the median is the mean of the 5th and 6th values; in the ordered list 5, 6, 6, 7, 7, 7, 8, 8, 9, 10 those are both 7, so the median is $(7+7)\div2=7$ marks. (c) Range $=10-5=5$ marks. (d) The sum is 73, so the mean is $73\div10=7.3$ marks. (The mean sits a little above the median, pulled up by the top mark of 10.)

Q4. The shop wants the size that sells most often — the most common value — which is the mode, **C**. The mean of the sizes need not be a size anyone wears, and the median only names the middle of the ordered list.

Q5. (a) Ordered, the middle (4th) value is 6: median 6 kg. The sum is 76, so the mean is $76 \div 7 \approx 10.9$ kg. (b) 40 kg — the next-largest mass is only 8 kg. (c) Without the 40 kg parcel the sum is 36 over six values: mean $36 \div 6 = 6$ kg. The mean fell by almost 5 kg while the median stayed 6 kg, so the outlier changed the mean more.

Q6. (a) Eggs are counted whole numbers: discrete, acquired by counting. (b) Mass is measured and can take any value within a range: continuous, acquired by measuring (a scale). (c) Temperature is measured: continuous, and yesterday's value is best found in a published source — the Bureau of Meteorology.

Q7. Ordered: 2.6, 5.0, 6.1, 6.8, 7.3, 7.6, 8.1, 8.2, 8.9, 9.0, 9.4, 9.5, 9.7, 9.8, 10.1, 10.3, 11.5, 11.6. Range $= 11.6 - 2.6 = 9.0$ °C. With 18 values the median is the mean of the 9th and 10th, $\left(8.9 + 9.0\right) \div 2 = 8.95$ °C. The sum is 151.5, so the mean is $151.5 \div 18 \approx 8.4$ °C. No value repeats, so there is no mode. The mean and median are close — both say a typical night was about 8–9 °C — with the one very cold night (2.6 °C) pulling the mean slightly below the median.

Q8. Set A: the sum is 40, mean $40 \div 5 = 8$; median (3rd value) 8; range $13 - 3 = 10$. Set B: the sum is 40, mean $40 \div 5 = 8$; median (3rd value) 8; range $13 - 3 = 10$. All three numbers agree, yet the sets differ — A holds three 8s and sits near the centre, B spreads evenly from end to end. Summarising with numbers alone loses that: the same summary can come from different data.

Q9. (a) The sum is 5600, so the mean is $5600 \div 7 = 800$ (thousand dollars); the median is the 4th value, 560 (thousand dollars). (b) \$800 000 is the mean. (c) The median: six of the seven houses sold between \$500 000 and \$600 000, so 560 thousand names a typical house; the one house at 2300 thousand drags the mean up to a price no house actually sold near.

Q10. (a) The values are 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 8. The tallest stack is 4: mode 4 people. The 8th value is 4: median 4 people. Range $= 8 - 2 = 6$ people. The sum is 66, so the mean is $66 \div 15 = 4.4$ people. (b) The median (or the mode), 4: most homes hold 3–6 people, and the single home with 8 pulls the mean above the middle of the data.

Q11. (a) Five values with mean 10 must sum to $5 \times 10 = 50$; the four given values sum to 42, so the fifth is $50 - 42 = 8$. (b) Ordered: 8, 8, 9, 12, 13 — median 9; mode 8 (the only repeat).

Q12. The sum is 532, so the mean is $532 \div 10 = 53.2$ sandwiches; the median is the mean of the 5th and 6th values, $\left(46 + 47\right) \div 2 = 46.5$ sandwiches. The 118 is an outlier — the next biggest day sold 52 — and it drags the mean above *every* ordinary day. To plan an ordinary day, the median (about 46–47) is the useful number.

Q13. Example answer 2, 4, 6, 6, 10. Check: the middle (3rd) value is 6, so the median is 6; 6 appears twice and no other value repeats, so the mode is 6; range $= 10 - 2 = 8$. Any set with a 6 in the middle, 6 repeated most often, and ends 8 apart works.

Q14. (a) The length of each student's best standing long jump, measured in cm. (b) Continuous — a jump length is measured and can take any value within a range. (c) Acquire the data by measuring: the same marked track for both year levels, each student's best of two or three jumps recorded in cm. (d) The median of each group is the number to compare; an outlier (one very long jump) would drag that group's mean towards it but would not move its median.

Q15. The sum is 107, so the mean is $107 \div 9 \approx 11.9$ years; the median is the 5th value, 9; the mode is 9 (three times); range $= 34 - 7 = 27$ years. The median (or the mode), 9, best describes a typical person: eight of the nine are students aged 7–11, and the coach's 34 drags the mean above every student's age.

Q16. (a) Five values with mean 8 sum to 40; the four given values sum to 31, so the fifth is 9. (b) Replacing 10 by 20 adds 10 to the sum, making it 50, so the new mean is $50 \div 5 = 10$ — no need to add everything again. (c) Before: 5, 7, 9, 9, 10, median 9; after: 5, 7, 9, 9, 20, median 9. The median only looks at the middle of the ordered list, and growing the largest value does not move the middle; the mean adds every value, so a bigger sum means a bigger mean.

Dot plots and stem-and-leaf plots: shape, centre and spread

Content description `VC2M7ST02`: create different types of displays of numerical data, including dot plots and stem-and-leaf plots, using software where appropriate; describe and compare the distribution of data, commenting on the shape, centre and spread including outliers and determining the range, median, mean and mode

Achievement standard (extract): Students plan and conduct statistical investigations involving discrete and continuous numerical data, using appropriate displays.

Theory

Every value in the picture. In 6A we met numerical data (discrete and continuous) and the four numbers that summarise a data set: the range, median, mean and mode. A summary number is useful but it is not a picture — two very different data sets can share the same median. A **display** draws every value at once, so the whole set is visible: where the values bunch, where they thin out, and whether any value sits far away from the rest. Two displays do this job at Level 7: the **dot plot** and the **stem-and-leaf plot**.

Dot plots. A dot plot is a number line with one dot drawn above each value, one dot per value; equal values stack upwards. Figure 1 shows the hours of sleep of 20 students the night before the sports carnival, as collected by the class.

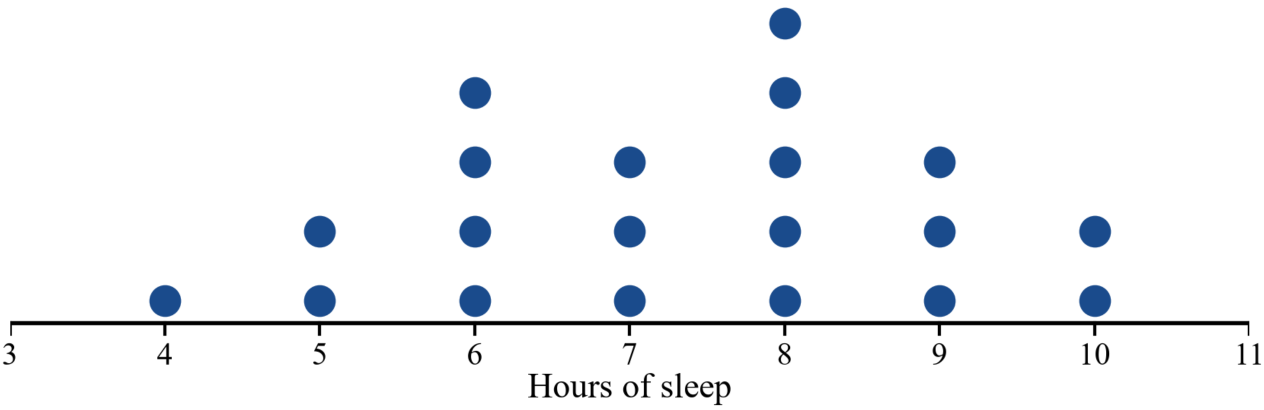

Figure 1. Dot plot of hours of sleep for 20 students: one dot above 4, two above 5, four above 6, three above 7, five above 8, three above 9 and two above 10.

Everything can be read straight off the picture. The values run from 4 to 10 hours, so the range is 6 hours; the tallest stack is at 8, so the mode is 8 hours; and the middle of the 20 values — the 10th and 11th, which are 7 and 8 — gives a median of 7.5 hours. Nothing is lost: every student's value is still there in the dots.

Stem-and-leaf plots. A stem-and-leaf plot splits each value into a **stem** (the leading digit or digits) and a **leaf** (the final digit). The stems stand in a column, smallest at the top, and each value's leaf is written out beside its stem. A **key** tells the reader how to put a stem and a leaf back into a value. For the eight values 12, 15, 18, 21, 21, 21, 23 and 26:

Stem	Leaf
1	2 5 8
2	1 1 1 3 6

Key: 1|2 means 12.

Writing the leaves in order gives an **ordered** stem-and-leaf plot, which lists the whole data set in ascending order — so the median and the range can be read off at once. Like a dot plot turned on its side, every value is still visible, because the leaves *are* the data. Two extensions are useful. If most of the leaves crowd onto one or two stems, **split stems** — writing each stem twice, once for leaves 0–4 and once for leaves 5–9 — spread the picture out. And to compare two data sets, a **back-to-back stem-and-leaf plot** shares one stem column between them, with one set's leaves on the right as usual and the other's on the left, read outwards from the stem.

Shape, centre and spread. The overall picture made by the dots or leaves is called the **distribution** of the data. To describe a distribution, comment on its **shape**, its **centre** (the median, mean or mode) and its **spread** (the range, and where the values cluster or leave gaps). The shape words used at Level 7 are:

- **symmetric** — the two sides look like mirror images of each other;
- **positive skew** — a tail of higher values stretches out to the right;
- **negative skew** — a tail of lower values stretches out to the left;
- **bi-modal** — two separate peaks, usually two **clusters** with a **gap** between them.

Figure 2 shows one of each shape, with the mean and median marked on every panel.

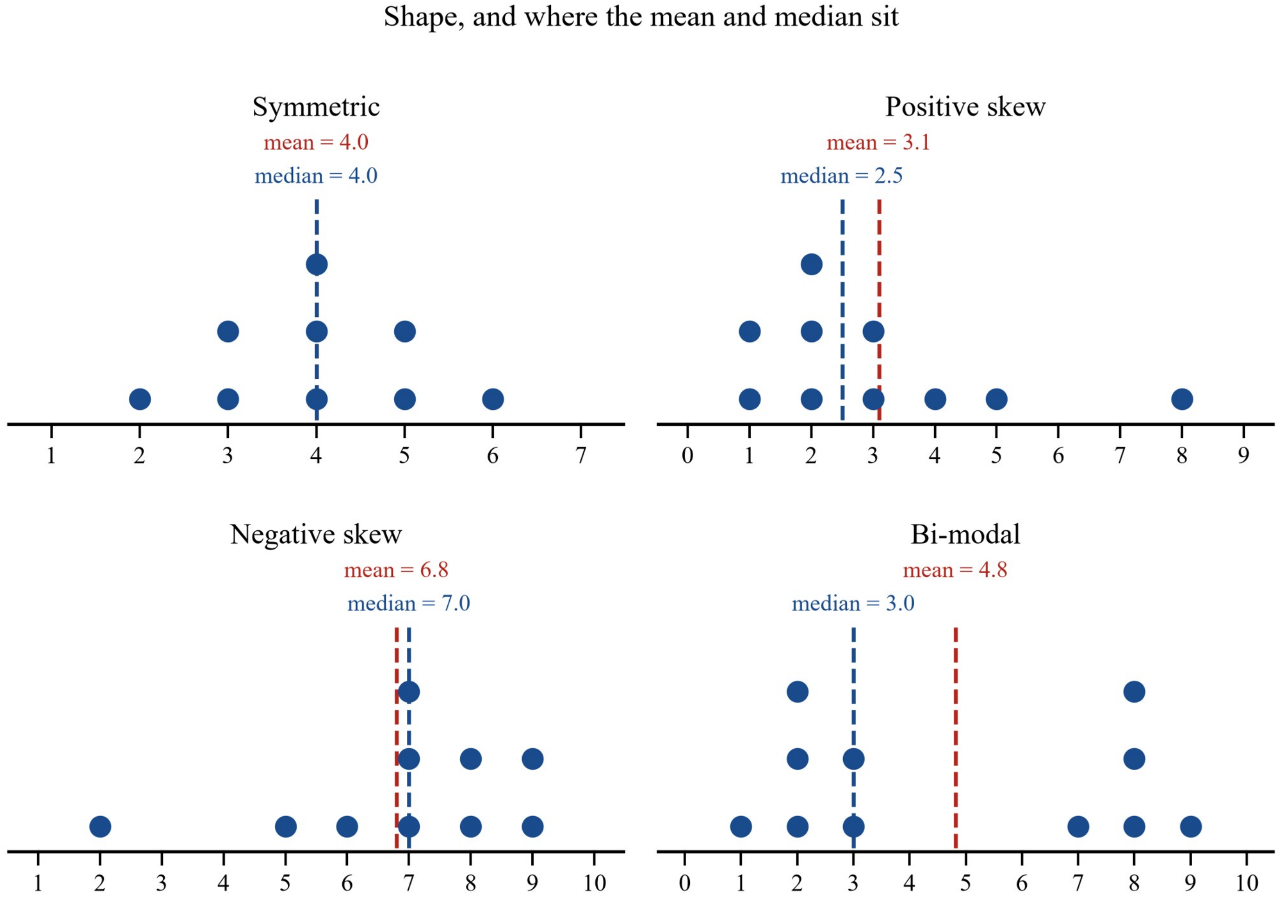

Figure 2. Four dot plots labelled symmetric, positive skew, negative skew and bi-modal, each with dashed lines marking its computed mean and median.

The features of a display connect straight to the summary numbers: the tallest stack or busiest leaf row is the **mode**, the two ends of the data give the **range**, and the middle value is the **median**. The shape also tells you how the mean and median sit. In a symmetric display they are close together near the middle. In a skewed display the **mean is pulled towards the tail**, so it sits away from the median. In a bi-modal display the mean and the median can both fall in the gap between the peaks, where there is little or no data, so neither is a good "typical" value.

Outliers. An **outlier** (6A) is a value far away from the rest of the data. The mean adds up every value, so one outlier pulls the mean towards it; the median only ever looks at the middle, so it does not move much. When a data set has an outlier, the median usually gives the truer centre — and working out the mean with and without the outlier shows the size of the pull.

Drawing displays with software. A spreadsheet or statistics program will draw dot plots and stem-and-leaf plots from a column of data in seconds, and it is the sensible choice once a data set gets large. Building a few displays by hand first is what teaches you to read them: the software draws the picture, but describing its shape, centre and spread is still your job.

The words of this subtopic, with synonyms you may meet:

Term	Meaning here	Synonyms and related words
dot plot	number line with one dot per value, equal values stacked	dot chart
stem-and-leaf plot	display splitting each value into a stem (leading digits) and a leaf (final digit)	stem plot
key	the line that tells how to read a stem and leaf back into a value	—
back-to-back stem-and-leaf plot	two data sets sharing one stem column, leaves on both sides	—
split stems	writing each stem twice (leaves 0–4, then 5–9) to spread a crowded display	—
distribution	the overall picture a display shows	the shape of the data
symmetric	the two halves look like mirror images	balanced
positive skew / negative skew	a tail stretching to the high side / the low side	skewed right / skewed left
bi-modal	two separate peaks	two clusters
cluster, gap	values bunched close together; a space holding no values	bunch, hole
outlier, range, median, mean, mode	as defined in 6A	—

Worked examples

Example 1 — scaffolded

The table gives the mean monthly maximum temperature at Melbourne Airport for each month of the year. Draw a dot plot of the twelve values, then find the range, median, mean and mode, and comment on the shape.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean max. temp. (°C)	26.7	26.7	24.3	20.4	16.7	13.7	13.2	14.6	16.9	19.6	22.1	24.7

Source: Bureau of Meteorology, "Climate statistics for Australian locations", station 086282 (Melbourne Airport), monthly means 1970–2026, generated 12 August 2026; retrieved 20 August 2026.

Working	Comment
Order the values: 13.2, 13.7, 14.6, 16.7, 16.9, 19.6, 20.4, 22.1, 24.3, 24.7, 26.7, 26.7.	Ordering first makes every later step easier.
Draw a number line from 13 to 27 and one dot above each value, stacking equal values (Figure 3).	Twelve months, so twelve dots.
Range $= 26.7 - 13.2 = 13.5$ °C.	Largest value minus smallest value.
Median: $n = 12$, so average the 6th and 7th values: $(19.6 + 20.4) \div 2 = 20.0$ °C.	An even number of values puts the median halfway between the middle two.

Working	Comment
Mean: the sum is 239.6, and $239.6 \div 12 \approx 20.0$ °C.	The mean uses every value.
Mode $= 26.7$ °C, the only value that repeats.	The one stack of two dots in Figure 3.
Shape: two clusters — the cool months (13–17 °C) and the warm months (24–27 °C) — with thinner ground between them; the mean and the median, both about 20 °C, sit in that gap.	A dot plot shows this at a glance. Here the mode, not the mean or median, names an actual cluster of months.

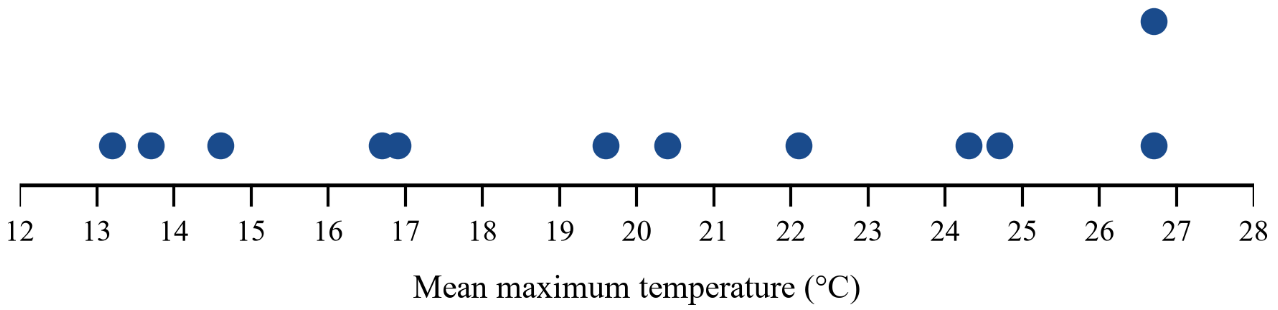

Figure 3. Dot plot of the twelve mean monthly maximum temperatures at Melbourne Airport, from 13.2 °C to 26.7 °C, with the two 26.7 °C values stacked and thinner ground between the cool and warm months.

Example 2 — scaffolded

A Year 7 class measured the height (cm) of each student in the room:

130, 134, 137, 141, 143, 145, 146, 148, 149, 152, 152, 155, 157, 159, 160, 162, 165, 168, 171, 174

Draw an ordered stem-and-leaf plot of the data, then use it to find the median, mode, mean and range, and describe the shape.

Working	Comment
The stems are 13, 14, 15, 16 and 17 (the hundreds and tens digits); write each height's ones digit as the leaf beside its stem, then order each row (Figure 4).	The key is 13|0 means 130 cm.
Median: with $n = 20$, average the 10th and 11th values. Counting down the leaves, both are 152, so the median is 152 cm.	The ordered plot lists the data in ascending order, so counting leaves is counting up the list.
Mode: no value repeats except 152, which appears twice, so the mode is 152 cm.	The busiest row is stem 14 (six leaves), but the mode is a repeated *value*, not a busy stem.
Mean: the sum is 3048, and $3048 \div 20 = 152.4$ cm.	The mean lands close to the median.
Range $= 174 - 130 = 44$ cm.	Read straight off the first and last leaves.
Shape: the leaves thin out at much the same rate above and below the 150s, and the mean (152.4 cm) sits close to the median (152 cm) — the display is roughly symmetric.	Compare the symmetric panel of Figure 2.

Heights of 20 Year 7 students (cm)

Stem	Leaf
13	0 4 7
14	1 3 5 6 8 9
15	2 2 5 7 9
16	0 2 5 8
17	1 4

Key: 13|0 means 130 cm

Figure 4. Ordered stem-and-leaf plot of the heights of 20 Year 7 students, stems 13 to 17, key 13|0 means 130 cm.

Example 3 — unscaffolded

Two paper-plane designs were each thrown 12 times by the same group of students. The flight distances (cm) were:

Design A: 283, 291, 305, 310, 318, 322, 327, 334, 341, 356, 360, 372 Design B: 245, 258, 264, 271, 279, 288, 297, 305, 312, 318, 325, 344

The data are drawn as a back-to-back stem-and-leaf plot in Figure 5. Use the display to compare the centre and the spread of the two designs, and state which design flies further.

Flight distances of two paper-plane designs (cm)

Design B	Stem	Design A
5	24	
8	25	
4	26	
1 9	27	
8	28	3
7	29	1
5	30	5
2 8	31	0 8
5	32	2 7
	33	4
4	34	1
	35	6
	36	0
	37	2

Key: B side 9|27 means 279 cm; A side 28|3 means 283 cm

Figure 5. Back-to-back stem-and-leaf plot of the flight distances of designs A and B, stems 24 to 37; the B leaves read outwards from the stem on the left.

Reading from the display, design A has median $(322+327)\div 2=324.5$ cm and mean $3919\div 12\approx 326.6$ cm, while design B has median $(288+297)\div 2=292.5$ cm and mean $3506\div 12\approx 292.2$ cm. For the spread, A runs from 283 to 372 cm (range 89 cm) and B from 245 to 344 cm (range 99 cm). Both means and both medians agree that design A flies further — about 34 cm further on average — while B's throws are a little more spread out. Both distributions look roughly symmetric, with no outliers.

$\therefore$ design A flies further (centre about 325 cm against about 292 cm); design B's results are slightly more spread (range 99 cm against 89 cm).

Example 4 — unscaffolded

Fifteen students estimated the length of their classroom (m):

8, 9, 9, 10, 10, 10, 11, 11, 12, 12, 12, 12, 13, 14, 25

The dot plot is shown in Figure 6, with the mean and median marked. Find the mean and the median, then find them again without the outlier, and explain which centre gives the truer picture of the class's estimates.

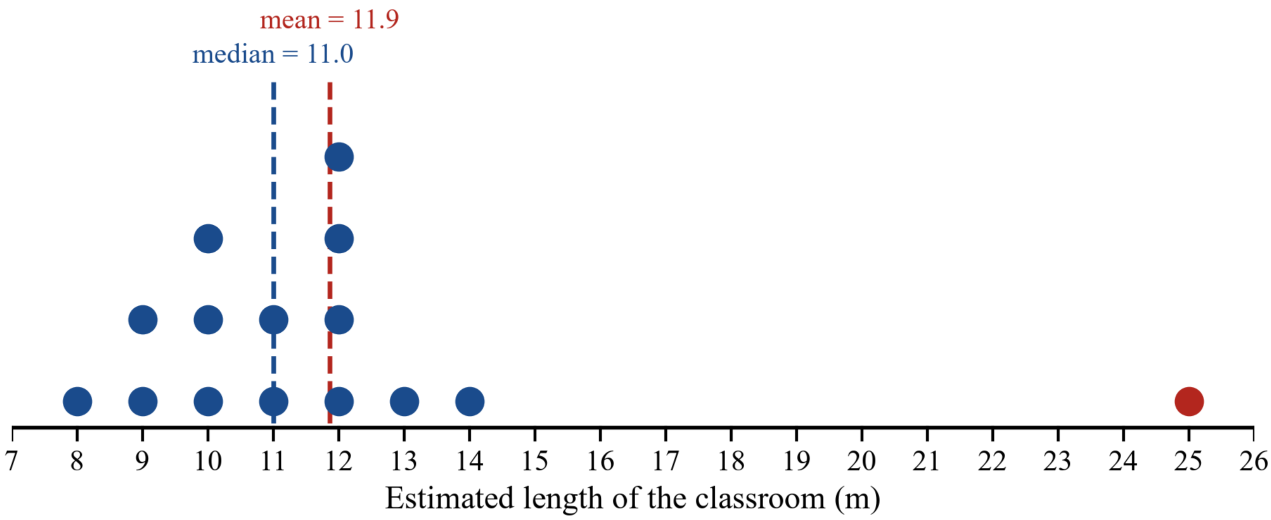

Figure 6. Dot plot of the fifteen estimates with the 25 m outlier drawn in red and dashed lines at the mean (11.9 m) and the median (11 m).

With every value included, the sum is 178, so the mean is $178 \div 15 \approx 11.9$ m; the middle (8th) value is 11, so the median is 11 m; and the range is $25 - 8 = 17$ m. Leaving out the 25 m estimate, the sum is 153 over 14 values, so the mean drops to $153 \div 14 \approx 10.9$ m, while the median stays 11 m and the range shrinks to $14 - 8 = 6$ m. One wrong guess moved the mean by a whole metre and did not move the median, because the mean adds up every value and the median only looks at the middle.

∴ with the outlier: mean ≈ 11.9 m, median 11 m; without it: mean ≈ 10.9 m, median 11 m. The median is the truer centre here — most estimates sat between 8 and 14 m.

Practice questions

Scaffolded

Q1. Figure 1 shows the hours of sleep of 20 students. (a) How many dots are on the plot, and why? (b) Write down the mode. (c) Find the median. (d) Find the range. (e) The mean is 7.3 hours. Is the mean smaller or larger than the median? (f) Which way does the tail point, and which shape word names it?

Q2. Twelve students counted the laps they each walked in a 20-minute charity walk:

21, 15, 18, 22, 16, 21, 19, 24, 17, 21, 16, 23

(a) Order the data from smallest to largest.
(b) Draw an ordered stem-and-leaf plot, with a key.
(c) Use your plot to find the median and the mode.
(d) Find the range.

Q3. The shoe sizes of 15 students are:

4, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9, 9

Draw a dot plot of the data, then find the mean (to 1 decimal place), the median, the mode and the range. Which is larger, the mean or the median, and what does that say about the tail?

Q4. This stem-and-leaf plot shows the marks out of 60 of 10 students on a science quiz.

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

Key: 3|1 means 31 marks.

(a) List the ten marks in ascending order.
(b) Find the median, the mode and the range.

Q5. A soccer team scored these goals in its 12 matches of the season:

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

(a) Which display — a dot plot or a stem-and-leaf plot — suits these data better, and why?
(b) Draw your chosen display.
(c) Find the median, the mode, the mean and the range.
(d) Comment on the shape.

Q6. Name the shape word (symmetric, positive skew, negative skew or bi-modal) that matches each description. (a) The two halves of the display look like mirror images. (b) A tail of high values stretches out to the right. (c) Two clusters of values sit with a gap between them. (d) A tail of low values stretches out to the left.

Unscaffolded

Q7. The table gives the mean monthly rainfall at Melbourne Airport for each month of the year.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean rainfall (mm)	42.4	38.6	36.9	45.3	39.7	39.7	35.0	43.6	45.2	55.7	62.4	51.0

Source: Bureau of Meteorology, "Climate statistics for Australian locations", station 086282 (Melbourne Airport), monthly means 1970–2026, generated 12 August 2026; retrieved 20 August 2026.

Draw a dot plot of the twelve values, then find the median, the mean (to 1 decimal place), the mode and the range, and describe the shape of the distribution.

Q8. The walking times (minutes) from home to school for 14 students were:

5, 8, 10, 12, 12, 12, 15, 18, 20, 22, 25, 25, 30, 35

Draw an ordered stem-and-leaf plot of the data, then find the median, the mode, the range and the mean (to 1 decimal place), and comment on the shape.

Q9. In Example 3, design A's longest throw of 372 cm was remeasured and found to be 380 cm. Without adding up the values again, state what happens to: (a) the range; (b) the median; (c) the mean.

Q10. Figure 7 shows the quiz marks out of 40 for 15 students, but one leaf is hidden by a question mark. The median of the 15 marks is 26. Find the hidden leaf, and find the range.

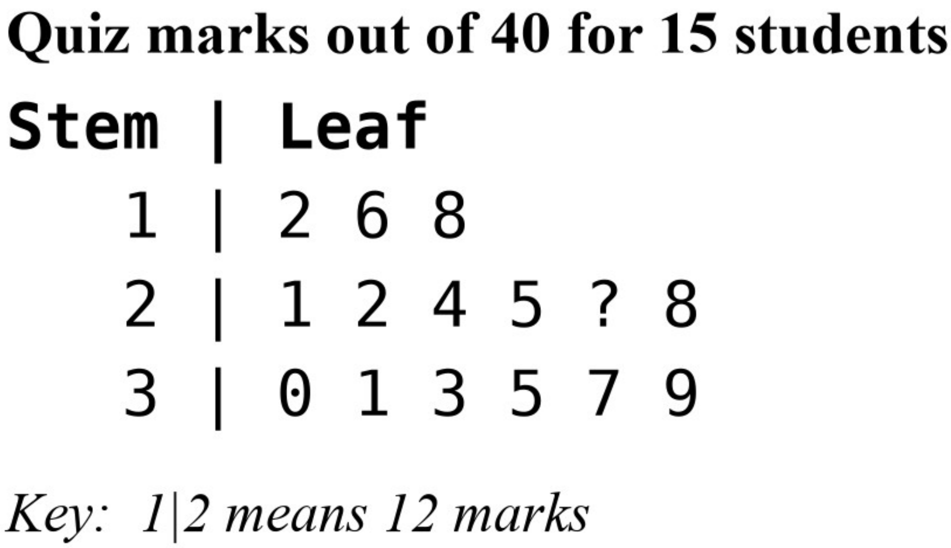

Figure 7. Stem-and-leaf plot of quiz marks out of 40 for 15 students, stems 1 to 3, with one leaf in the stem-2 row hidden by a question mark; key 1|2 means 12 marks.

Q11. Figure 8 shows the number of books read in Term 2 by the 13 students of class 7A and the 13 students of class 7B. Compare the two classes using centre and spread: give each class's median, mean (to 1 decimal place) and range, and comment on the shape of each display.

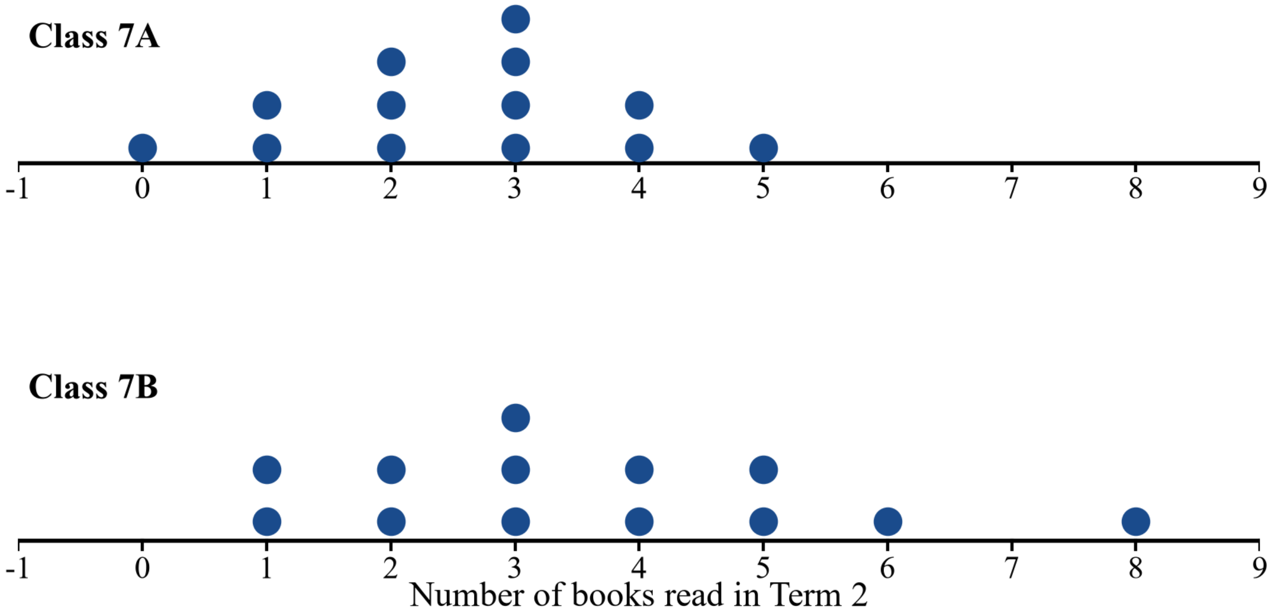

Figure 8. Two dot plots on the same number line, one above the other: books read in Term 2 by class 7A and by class 7B, values from 0 to 8.

Q12. The masses (kg) of ten parcels on a desk are:

2, 3, 3, 4, 4, 4, 5, 5, 6, 24

Find the mean and the median of the ten masses, then find them again without the 24 kg parcel. Which centre was changed more by the outlier?

Q13. Four dot plots are drawn in Figure 9, labelled A to D. Name the shape of each distribution, using the words symmetric, positive skew, negative skew and bi-modal.

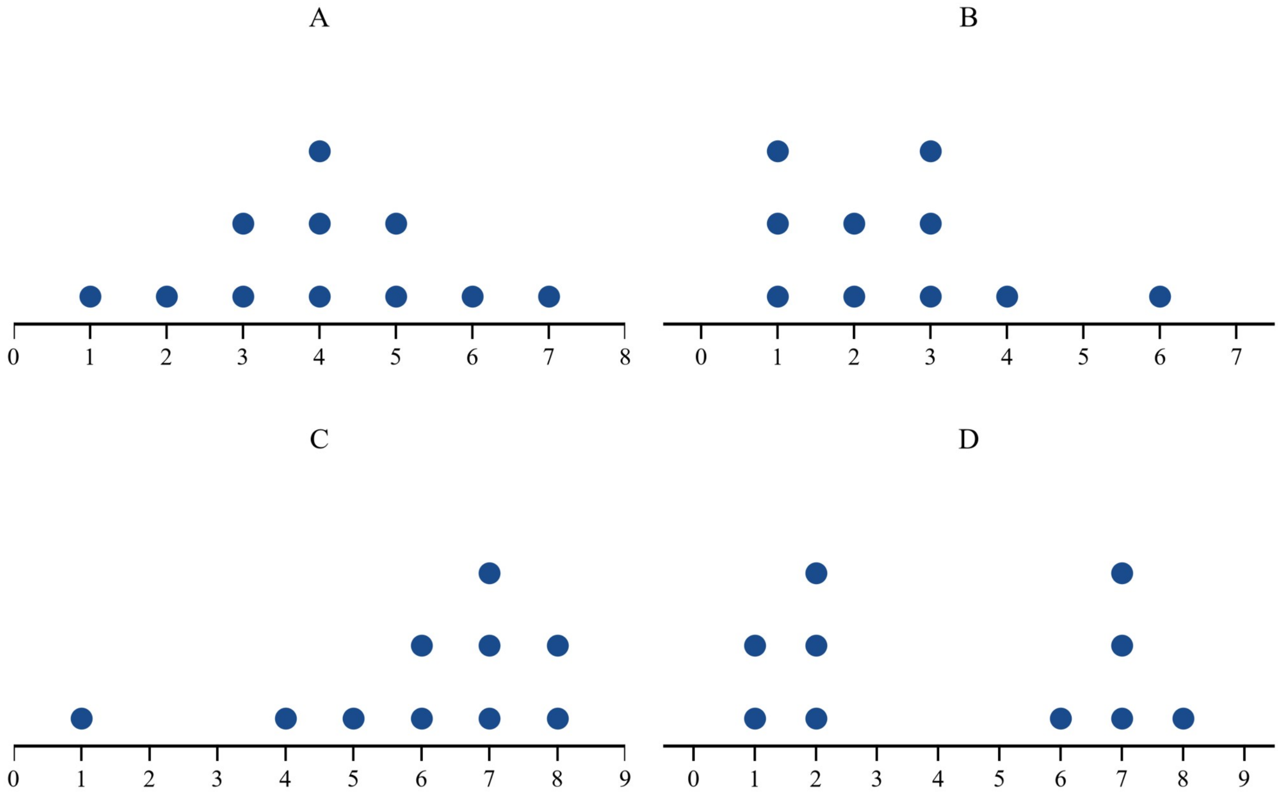

Figure 9. Four unlabelled-shape dot plots A to D: A peaks in the middle with equal tails, B has a tail to the high side, C has a tail to the low side, and D has two separate peaks.

Q14. Twenty players' best scores (points) in a video game were:

12, 15, 18, 21, 22, 25, 27, 29, 30, 33, 35, 36, 38, 41, 44, 45, 48, 52, 58, 72

(a) Explain why a stem-and-leaf plot is a good display for these data.
(b) Find the median and the mean.
(c) One value is an outlier. Identify it and describe its effect on the mean.

Q15. For each job, name the display that suits it best (dot plot, stem-and-leaf plot, or back-to-back stem-and-leaf plot) and explain your choice in one sentence. (a) Show the number of siblings of each of 25 students (values 0 to 6). (b) Compare the marks of two classes on the same spelling test. (c) Show the masses (kg), all two-digit values, of 40 suitcases at an airport check-in desk.

Q16. The marks out of 50 of 16 students on a maths quiz were:

21, 23, 25, 27, 28, 30, 32, 34, 36, 38, 40, 41, 43, 45, 47, 49

An ordinary stem-and-leaf plot puts all 16 leaves on only three stems. Draw the plot using split stems (each stem written twice, for leaves 0–4 then 5–9), then find the median and the range.

Q17. Two data sets are:

A: 5, 7, 8, 9, 11 B: 1, 6, 8, 10, 15

Find the mean, median and range of each set. What do the two sets share, and how do they differ?

Q18. A netball team's scores (goals) across 18 matches were:

4, 6, 7, 9, 11, 12, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 26, 38

Draw a display of the data, find the median, the mode, the range and the mean (to 1 decimal place), and identify the outlier. Which centre — the mean or the median — would you report as the team's "typical" score, and why?

Answers

Q1. (a) 20 — one dot per student. (b) 8 hours. (c) 7.5 hours. (d) 6 hours. (e) smaller (7.3 < 7.5). (f) towards the low values; negative skew.

Q2. (a) 15, 16, 16, 17, 18, 19, 21, 21, 21, 22, 23, 24. (b) Stems 1 and 2: 1 | 5 6 6 7 8 9 and 2 | 1 1 1 2 3 4, key 1|5 means 15 laps. (c) Median 20 laps; mode 21 laps. (d) 9 laps.

Q3. Mean ≈ 6.5; median 6; mode 6; range 5. The mean is larger than the median: the tail of larger sizes (8 and 9) pulls the mean up — a slight positive skew.

Q4. (a) 31, 34, 37, 40, 42, 45, 45, 49, 53, 56. (b) Median 43.5 marks; mode 45 marks; range 25 marks.

Q5. (a) A dot plot — the values are single digits that repeat, so the stacks do the work. (b) Stacks of 1, 2, 3, 3, 3, 1, 2, 2, 1 dots above 0 to 8. (c) Median 3; mode 3; mean 3.5; range 8. (d) A tail of higher scores to the right: positive skew (the mean, 3.5, sits above the median, 3).

Q6. (a) symmetric. (b) positive skew. (c) bi-modal. (d) negative skew.

Q7. Median 43.0 mm; mean ≈ 44.6 mm; mode 39.7 mm; range 27.4 mm. Most months sit between 35 and 46 mm with a tail of wetter months (51.0, 55.7, 62.4): positive skew, and the mean sits above the median.

Q8. Median 16.5 min; mode 12 min; range 30 min; mean ≈ 17.8 min. A tail of longer times to the right: positive skew, with the mean (17.8) pulled above the median (16.5).

Q9. (a) It grows to $380-283=97$ cm. (b) It stays 324.5 cm — the middle two values are unchanged. (c) It grows (the sum is bigger over the same 12 throws).

Q10. The hidden leaf is 6 (the mark 26); the range is $39-12=27$ marks.

Q11. 7A: median 3, mean ≈ 2.5, range 5. 7B: median 3, mean ≈ 3.6, range 7. The medians are equal but 7B's mean is higher, so 7B read more books overall; 7B is also more spread out, with a tail of big readers (positive skew).

Q12. With the outlier: mean 6.0 kg, median 4 kg. Without it: mean 4.0 kg, median 4 kg. The mean was changed more (it fell by 2 kg; the median did not move).

Q13. A symmetric; B positive skew; C negative skew; D bi-modal.

Q14. (a) Twenty two-digit values: the stems 1 to 7 keep every value visible and show the shape at a glance. (b) Median 34 points; mean 35.05 points. (c) 72; it sits far above the rest and pulls the mean up above the median.

Q15. (a) Dot plot — small whole numbers with repeats stack well. (b) Back-to-back stem-and-leaf plot — it puts the two classes side by side for comparing. (c) Stem-and-leaf plot (or software) — two-digit values that would not stack well as a dot plot.

Q16. Rows: 2 (0–4) | 1 3; 2 (5–9) | 5 7 8; 3 (0–4) | 0 2 4; 3 (5–9) | 6 8; 4 (0–4) | 0 1 3; 4 (5–9) | 5 7 9. Median 35; range 28.

Q17. A: mean 8, median 8, range 6. B: mean 8, median 8, range 14. They share the same centre (mean and median both 8) but differ in spread: B is much more spread out.

Q18. Median 14.5; mode 12; range 34; mean ≈ 15.9; outlier 38. Report the median — the outlier pulls the mean up, so the median gives the truer "typical" score.

Fully-worked solutions

Q1. (a) One dot is drawn per value, so there are 20 dots, one for each student. (b) The tallest stack is at 8 (five dots): mode 8 hours. (c) With 20 values the median is the average of the 10th and 11th. Counting up: 1 + 2 + 4 = 7 values are 6 or less, so the 8th to 10th values are 7 and the 11th to 15th are 8; the median is $(7+8)\div 2=7.5$ hours. (d) Range $=10-4=6$ hours. (e) 7.3 < 7.5, so the mean is smaller than the median. (f) The tail of low values (the lone 4) points left: negative skew.

Q2. (a) 15, 16, 16, 17, 18, 19, 21, 21, 21, 22, 23, 24. (b) The stems are the tens digits, 1 and 2:

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

Key: 1|5 means 15 laps. (c) With 12 values the median is the average of the 6th and 7th, $(19+21)\div 2=20$ laps; 21 appears three times and no other value repeats, so the mode is 21 laps. (d) Range $=24-15=9$ laps.

Q3. The dot plot stacks four dots at 6 and three each at 5 and 7. Mean: the sum is 98, and $98\div 15\approx 6.5$. The median is the 8th value, 6. The mode is 6 (four dots). Range $=9-4=5$. The mean (6.5) is larger than the median (6): the few large sizes (8 and 9) form a tail to the right that pulls the mean up — a slight positive skew.

Q4. (a) Reading each leaf back with its stem: 31, 34, 37, 40, 42, 45, 45, 49, 53, 56. (b) With 10 values the median is the average of the 5th and 6th, $(42+45)\div 2=43.5$ marks. The mode is 45 (the only repeated leaf). Range $=56-31=25$ marks.

Q5. (a) The values are single digits, several of them repeated, so a dot plot suits them best: the stacks show the shape at once, and a stem-and-leaf plot would have only one stem. (b) Dots above each value: one at 0, two at 1, three at 2, three at 3, one at 4, two at 5, none at 6 or 7, one at 8. (c) The median is the average of the 6th and 7th values, $(3+3)\div 2=3$; the mode is 3; the mean is $42\div 12=3.5$; the range is $8-0=8$. (d) The higher scores form a tail to the right and the mean (3.5) sits above the median (3): positive skew.

Q6. (a) Mirror-image halves mean **symmetric**. (b) A tail of high values to the right is **positive skew**. (c) Two clusters with a gap between them is **bi-modal**. (d) A tail of low values to the left is **negative skew**.

Q7. Ordering the values: 35.0, 36.9, 38.6, 39.7, 39.7, 42.4, 43.6, 45.2, 45.3, 51.0, 55.7, 62.4. The median is the average of the 6th and 7th, $(42.4+43.6)\div 2=43.0$ mm. The sum is 535.5, so the mean is $535.5\div 12\approx 44.6$ mm. The mode is 39.7 mm (the only repeat). Range $=62.4-35.0=27.4$ mm. Ten of the twelve months sit between 35 and 46 mm while three wetter months (51.0, 55.7, 62.4) form a tail to the right: positive skew, matching the mean sitting above the median.

Q8. Stems 0 to 3: 0 | 5 8; 1 | 0 2 2 2 5 8; 2 | 0 2 5 5; 3 | 0 5, with key 0|5 means 5 minutes. The median is the average of the 7th and 8th values, $(15+18)\div 2=16.5$ min. The mode is 12 (three dots). Range $=35-5=30$ min. The sum is 249, so the mean is $249\div 14\approx 17.8$ min. The longer times form a tail to the right and the mean (17.8) sits above the median (16.5): positive skew.

Q9. (a) The range uses the largest value: it becomes $380-283=97$ cm instead of 89 cm, so it grows. (b) The median is the average of the 6th and 7th values (322 and 327), and changing the largest value does not move them: the median stays 324.5 cm. (c) The mean adds up every value; a bigger sum over the same 12 throws means the mean grows.

Q10. There are 15 marks, so the median is the 8th. The stem-1 row holds the 1st to 3rd values, so the 8th value is the 5th leaf in the stem-2 row — the hidden one. A median of 26 means that leaf is 6. The range does not need the hidden leaf at all: $39-12=27$ marks.

Q11. Class 7A: the median (7th value) is 3; the sum is 33, so the mean is $33\div 13\approx 2.5$; range $=5-0=5$. Class 7B: the median (7th value) is 3; the sum is 47, so the mean is $47\div 13\approx 3.6$; range $=8-1=7$. Both classes have the same median, but 7B's larger mean says it read more books overall, and its bigger range plus a tail of large readers (the 6 and the 8) shows more spread and a positive skew, against 7A's narrower, more symmetric display.

Q12. With all ten parcels: the sum is 60, mean $60\div 10=6.0$ kg; the median is the average of the 5th and 6th values, $(4+4)\div 2=4$ kg. Without the 24 kg parcel: the sum is 36 over 9 values, mean $36\div 9=4.0$ kg; the median (5th value) is still 4 kg. The outlier moved the mean by 2 kg and the median not at all, so the mean was changed more.

Q13. A: equal tails either side of a middle peak — symmetric. B: a tail stretching to the high side — positive skew. C: a tail stretching to the low side — negative skew. D: two separate peaks with a gap — bi-modal.

Q14. (a) The values are two-digit and there are twenty of them: stems 1 to 7 keep every value visible, order the data, and show the shape in one picture. (b) With 20 values the median is the average of the 10th and 11th, $(33+35)\div 2=34$ points. The sum is 701, so the mean is $701\div 20=35.05$ points. (c) The outlier is 72 — the next-highest score is only 58. It pulls the mean up above the median; without it the mean would sit closer to the median.

Q15. (a) Dot plot: the values are small whole numbers (0 to 6) with plenty of repeats, so the stacks show everything at once. (b) Back-to-back stem-and-leaf plot: it draws both classes against one stem column, so their centres and spreads can be compared directly. (c) Stem-and- leaf plot (or software): two-digit masses would not stack usefully on a dot plot, but stems and leaves keep every value and show the shape.

Q16. With split stems, each stem appears twice — first for leaves 0–4, then for leaves 5–9:

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

Key: 3 (5–9)|6 means 36. With 16 values the median is the average of the 8th and 9th, $(34+36)\div 2=35$ marks. Range = $49-21=28$ marks. The split stems spread the leaves so the shape is easier to see.

Q17. Set A: mean = $40\div 5=8$; median (3rd value) = 8; range = $11-5=6$. Set B: mean = $40\div 5=8$; median (3rd value) = 8; range = $15-1=14$. The two sets have exactly the same centre — mean and median both 8 — but very different spread: B's range (14) is more than double A's (6). Centre alone does not describe a data set.

Q18. An ordered stem-and-leaf plot (stems 0 to 3) or dot plot shows the data. With 18 values the median is the average of the 9th and 10th, $(14+15)\div 2=14.5$ goals. The mode is 12 (the only repeat). Range = $38-4=34$ goals. The sum is 287, so the mean is $287\div 18\approx 15.9$ goals. The 38 is an outlier — the next-highest score is 26 — and it pulls the mean (15.9) above the median (14.5). Report the median as the typical score: the one big score does not move it, while the mean is pulled up by it.

Planning and conducting a statistical investigation

Content description VC2M7ST03: plan and conduct statistical investigations for issues involving discrete and continuous numerical data, and data collected from primary and secondary sources; analyse and interpret distributions of data and report findings in terms of shape and summary statistics

Achievement standard (extract): Students plan and conduct statistical investigations involving discrete and continuous numerical data, using appropriate displays.

Theory

Statistical questions. Some questions are answered by looking up a fact: "What is the capital of Victoria?" needs no data at all. Other questions are answered by collecting data, and the data will *vary* from one person or thing to the next: "How do students in Year 7 travel to school?" needs data, and the answers differ (some walk, some ride, some catch a train). A question like that, answered with data that vary, is a **statistical question**. To answer one well, you run a **statistical investigation**: you collect the data, look at the whole set, and say what it shows.

The investigation cycle. Figure 1 shows the six stages of a statistical investigation. The loop is the important part: if your report does not really answer the question you asked, you go back and fix the plan or collect better data.

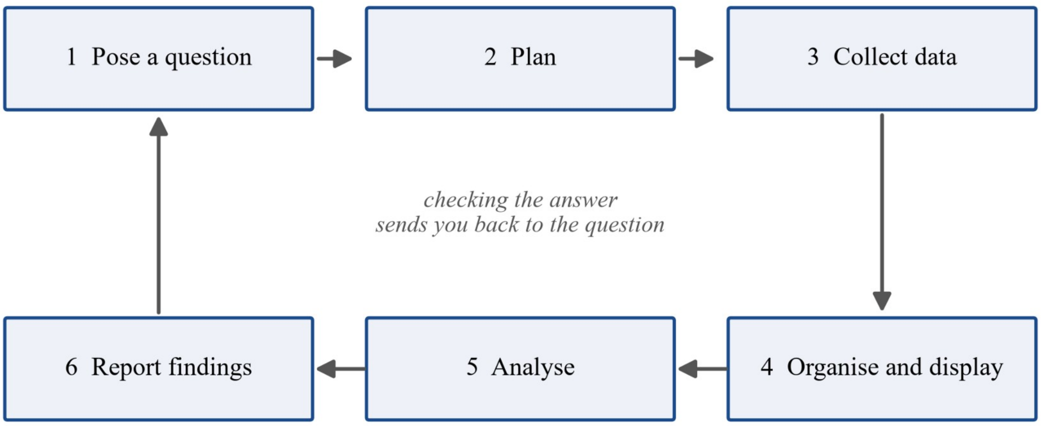

Figure 1. The six stages of a statistical investigation arranged in a loop: pose a question, plan, collect data, organise and display, analyse, report findings, with an arrow from the last stage back to the first.

The data: discrete and continuous. This subtopic uses **numerical** data, and subtopic 6A split numerical data into two kinds, shown again in Figure 2. **Discrete** data are counted and take whole-number values (number of pets, goals kicked in a game). **Continuous** data are measured and can take any value within a range (height in cm, time in seconds). The kind of variable decides how you measure it and which display suits it, so name the kind in your plan.

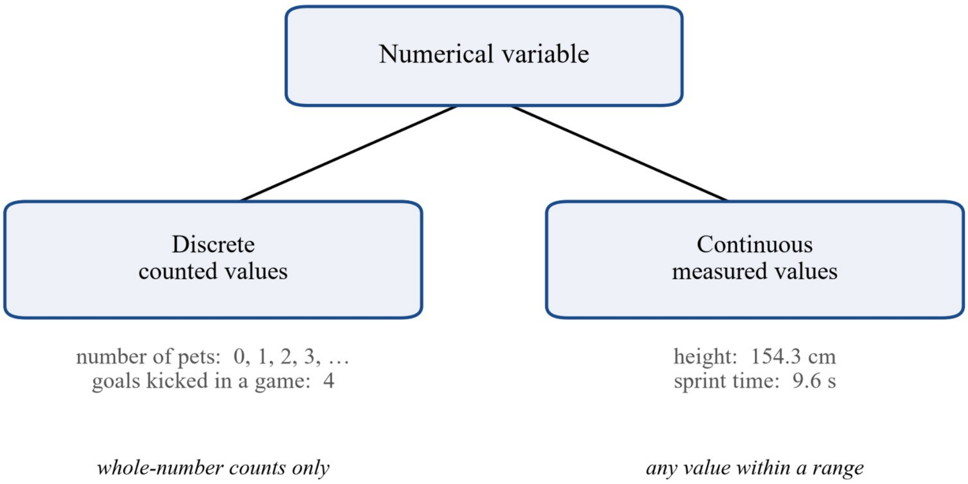

Figure 2. A tree diagram: a numerical variable splits into discrete, counted values (number of pets 0, 1, 2, 3, …; goals kicked in a game 4) and continuous, measured values (height 154.3 cm; sprint time 9.6 s).

Primary and secondary sources. There are two ways to get data. A **primary source** is one where *you* collect the data yourself — a survey of your class, your own measurements, your own experiment. A **secondary source** is one where someone else collected the data and you obtain it — from newspapers, the internet or the Australian Bureau of Statistics. Figure 3 shows both. Each has its place: primary data give you exactly the variable you want, while secondary data can cover far more than you could measure yourself — the ABS publishes data about the whole country, for example. At Level 7 this split between primary and secondary sources is the only one you need: write into your plan where the data will come from and why.

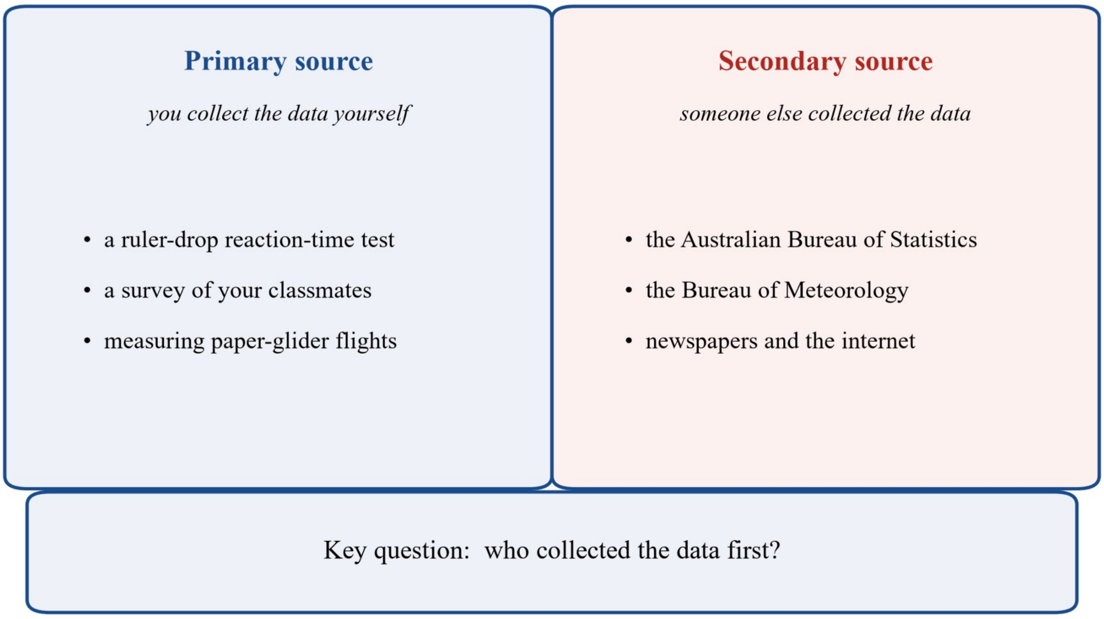

Figure 3. Two panels. Primary source — you collect the data yourself: a ruler-drop reaction-time test, a survey of your classmates, measuring paper-glider flights. Secondary source — someone else collected the data: the Australian Bureau of Statistics, the Bureau of Meteorology, newspapers and the internet. A strip below asks the key question: who collected the data first?

Planning. A good plan answers these questions before any data are collected:

- *What is the statistical question?* It must name **who or what** is studied and **what will be measured**.
- *What is the variable, and what units will it be measured in?* Name it as discrete or continuous.
- *Where will the data come from* — a primary or a secondary source?

- *How many observations?* More data usually give a clearer picture; plan as many as you can collect and check in the time.
- *How will the data be displayed?* At this level, a dot plot or a stem-and-leaf plot (subtopic 6B) is the usual choice.
- *If people are involved, is it fair and is it allowed?* Ask before you measure people, and measure everyone the same way.

Conducting. Carry out the plan exactly as written: the same tool, the same units and the same method every time, so the data can be compared in a fair way. Record every observation straight into a table as it is collected — data written on scraps of paper get lost or mixed up.

Analysing: shape, centre and spread. Once the data are displayed, describe the whole set the way subtopics 6A and 6B do. Figure 4 shows the four shape words you must know: **symmetric**, **positive skew**, **negative skew** and **bi-modal**. A **cluster** is a group of values sitting close together; a **gap** is a stretch of the axis with no values; an **outlier** is a value far away from the rest of the data. For centre, use the **median**, the **mean** and the **mode**, and for spread, the **range**. The shape tells you which centre to trust: in a symmetric set the mean and median are much the same, but in a skewed set, or a set with an outlier, the mean is pulled toward the tail while the median holds its place in the middle.

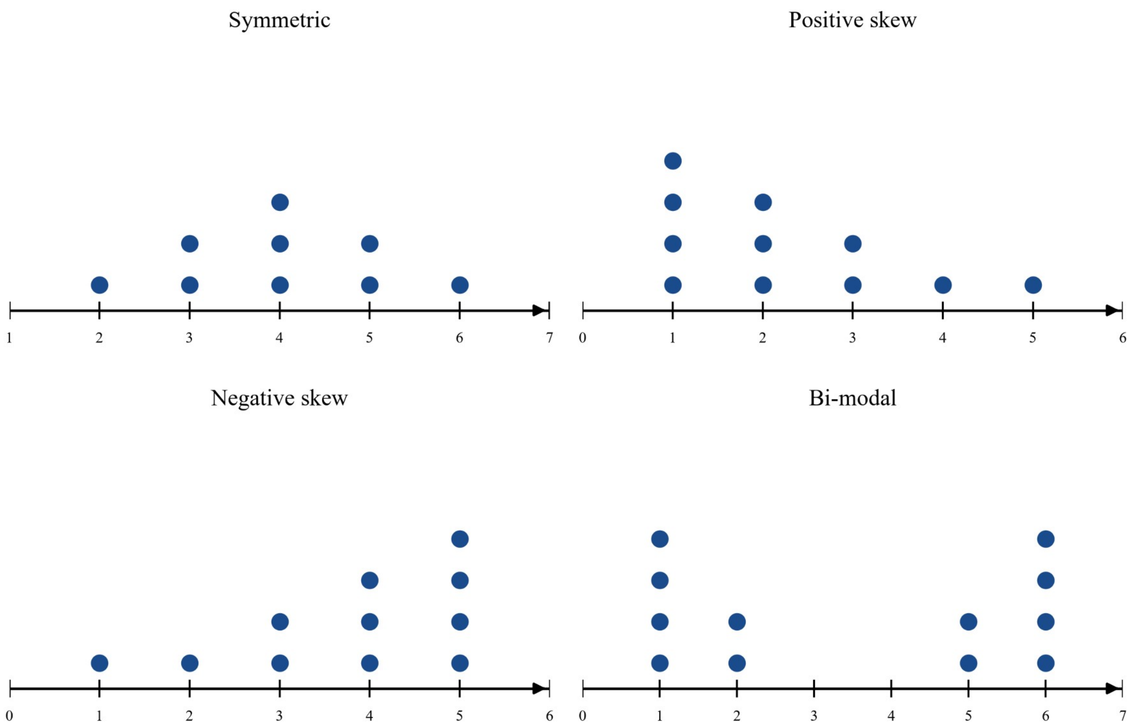

Figure 4. Four dot plots showing the four shape words. Symmetric: a peak in the middle with even tails. Positive skew: a tail stretching toward the larger values. Negative skew: a tail stretching toward the smaller values. Bi-modal: two separate peaks.

Reporting findings. The report answers the original question *in words*, using the shape and the summary statistics as the reasons. Useful starters:

- "The distribution of … was (symmetric / positively skewed / …), with a cluster around …"
- "The median was …, so a typical … was about …"
- "The range was …, which shows the data (were close together / spread out)."
- "The mean was (larger / smaller) than the median because …"

If the data came from a secondary source, name the source and the date it was collected or published — a finding is only as trustworthy as its source. These four habits — forming the question, choosing the plan, doing the mathematics accurately, and interpreting and reporting it in words — are exactly what Investigation task **I1** assesses.

The words of this subtopic, with synonyms you may meet:

Term	Meaning here	Synonyms and related words
statistical question	a question answered with data that vary	investigative question
variable	the thing that is counted or measured	what the data are about
discrete	counted; whole-number values	counts
continuous	measured; any value in a range	measurements
primary source	data you collect yourself	first-hand data
secondary source	data someone else collected	published data
distribution	the overall pattern of a set of data	the shape of the data
outlier	a value far from the rest of the data	unusual value, extreme value

Subtopic 6A owns the centre and spread words (range, median, mean, mode, outlier) and 6B owns the displays (dot plots and stem-and-leaf plots); this subtopic uses all of them inside the investigation cycle, and it is where the skills are assessed in task I1.

Worked examples

Example 1 — scaffolded

Plan a primary-source investigation for the question: "Do adults have faster reaction times than Year 7 students?" using the ruler-drop test of Figure 5.

How the test works (Figure 5): a 20 cm ruler is held at the top and dropped without warning. The other person waits with thumb and finger open at the 0 cm end and catches the ruler as fast as they can. The number on the ruler at the catch point is the **catch distance** — a shorter catch distance means a faster reaction time.

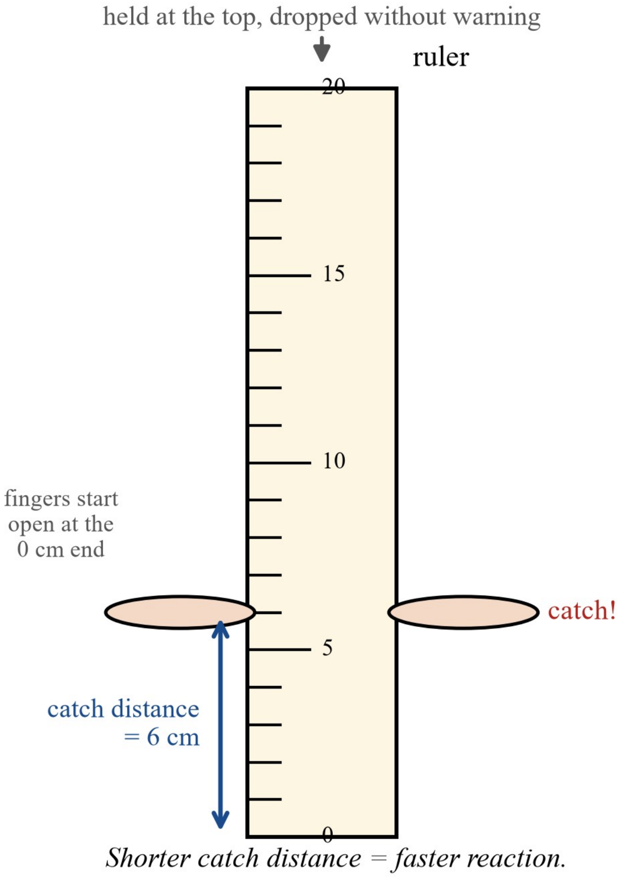

Figure 5. The ruler-drop test: a 20 cm ruler is held at the top and dropped without warning, and the catcher closes their fingers on it; the catch distance of 6 cm is read at the fingers, and a shorter catch distance means a faster reaction.

Working	Comment
Question: Do adults have faster reaction times than Year 7 students? Method: the ruler-drop test, the same for both groups.	A statistical question — catch distances will vary from person to person, so both groups must be tested the same way.
Variable: catch distance in cm. It is measured and can take any value in a range, so it is **continuous**.	Naming the kind of variable decides the display and the summary statistics.
Source: **primary** — measure 10 Year 7 students and 10 adults yourself; three trials each person, keeping the middle (median) of the three.	Three trials stop one lucky or unlucky catch from spoiling a person's result.
Display: a dot plot for each group, side by side (subtopic 6B).	Two matching displays make the groups easy to compare.
Analysis and report: compare the medians and ranges of the two groups; shorter distances mean faster reactions, so the group with the smaller median reacted faster.	The report answers the question in words, with the medians and ranges as the reasons.

∴ the plan can answer the question, because both groups are measured the same way and the report compares a centre and a spread for each group.

Example 2 — scaffolded

The table gives the estimated resident population of each Australian state and territory at 31 December 2025. Treat it as secondary data: display it, describe it, and report one finding.

State or territory	Population (thousands)
Northern Territory	267.5
Australian Capital Territory	487.2
Tasmania	579.1
South Australia	1 910.6
Western Australia	3 076.5
Queensland	5 712.1
Victoria	7 121.9
New South Wales	8 641.1

Source: Australian Bureau of Statistics, "National, state and territory population", December 2025; retrieved 20 August 2026. (The Australia total also includes the Other Territories, which are not listed.)

Working	Comment
Source: **secondary** — the ABS collected these counts; we obtain them from the ABS. Variable: population of a state or territory, in people. It is counted, so it is **discrete**.	Naming the source in the plan is part of the CD. A count of people is a whole number, even when the table rounds to thousands.
Display: a dot plot, Figure 6.	Eight values spread over a wide range — a dot plot shows the spread and the skew clearly.
Median: middle pair 1 910.6 and 3 076.5, so $(1910.6+3076.5)\div 2=2493.55$ thousand $=2\,493\,550$ people. Range: $8641.1-267.5=8373.6$ thousand $=8\,373\,600$ people. Mean: the eight values add to 27 796.0 thousand, and $27\,796.0\div 8=3474.5$ thousand $=3\,474\,500$ people.	With eight values the median is the mean of the 4th and 5th after ordering.
Shape: **positive skew** — five of the eight dots sit low, with a long tail toward the larger states; the mean (3.47 million) is larger than the median (2.49 million), just as	Compare the two dashed lines in Figure 6: the mean has been pulled toward the tail.

Working	Comment

positive skew predicts.

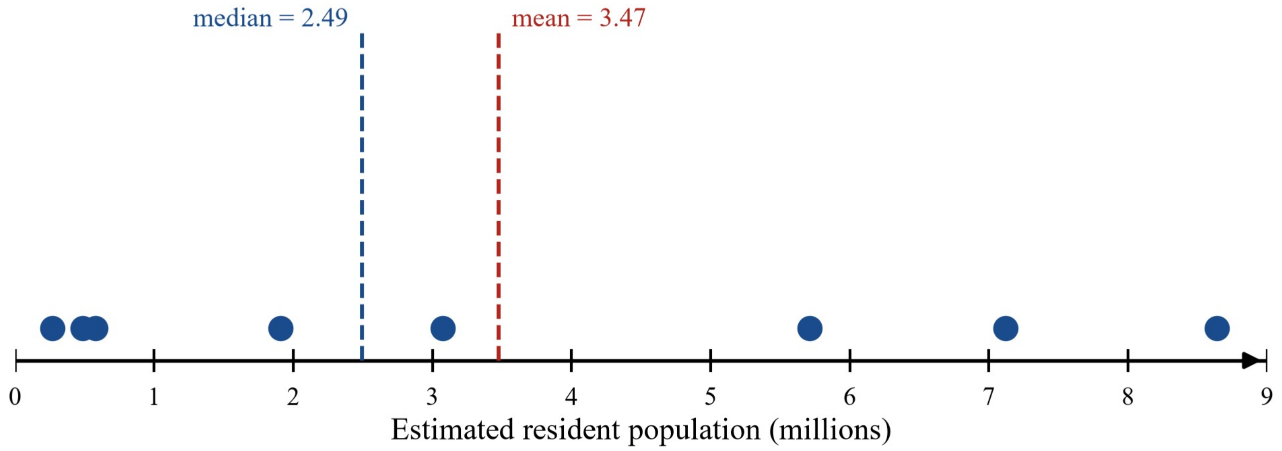

Figure 6. Dot plot of the estimated resident population of each Australian state and territory at 31 December 2025, in millions: dots at 0.27, 0.49, 0.58, 1.91, 3.08, 5.71, 7.12 and 8.64, with dashed lines at the median 2.49 and the mean 3.47.

∴ a typical state or territory had about 2.5 million people (the median), while the mean of about 3.5 million was pulled up by the two large states; the distribution is positively skewed, with five of the eight below the mean.

Example 3 — unscaffolded

The table gives the annual movement of the consumer price index (CPI) — the ABS measure of how prices change — for eight quarters. Describe the distribution (shape, centre, spread) and report one finding.

Quarter	Annual CPI movement (%)
Sep 2024	1.9
Dec 2024	2.4
Mar 2025	2.4
Jun 2025	2.9
Sep 2025	3.6
Dec 2025	3.8
Mar 2026	4.6
Jun 2026	3.8

Source: Australian Bureau of Statistics, "Consumer Price Index, Australia", June quarter 2026; retrieved 20 August 2026.

Ordering the values gives 1.9, 2.4, 2.4, 2.9, 3.6, 3.8, 3.8, 4.6. The median is the mean of the middle pair, $(2.9+3.6)\div 2=3.25$; the mean is $25.4\div 8=3.175$, about 3.18; the range is $4.6-1.9=2.7$. The dot plot (Figure 7) shows the striking feature: the values fall into two clusters, 1.9–2.9 and 3.6–4.6, with a gap between — the distribution is **bi-modal**.

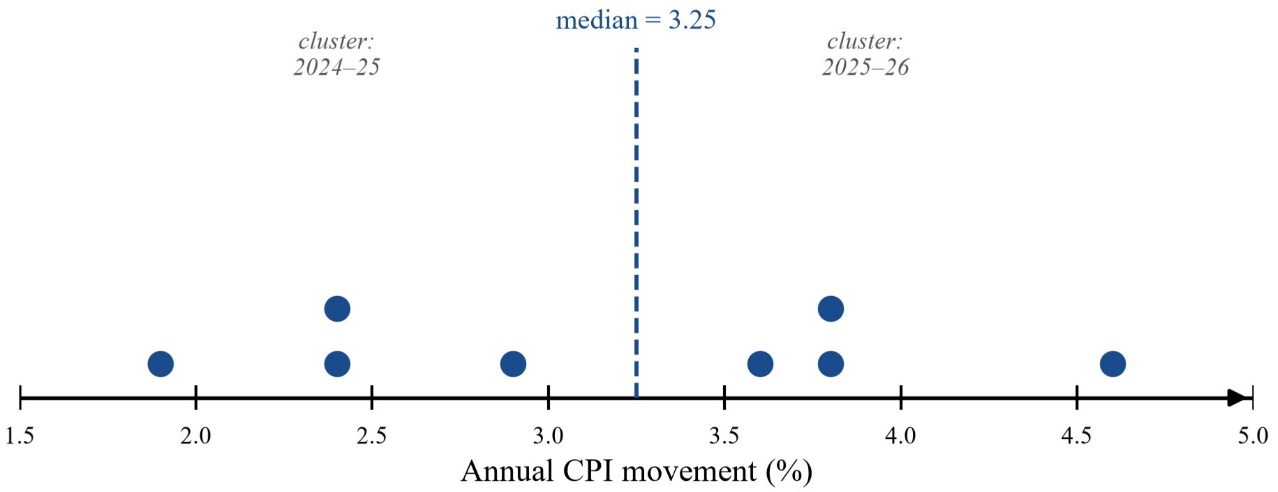

Figure 7. Dot plot of the annual CPI movement for the eight quarters September 2024 to June 2026, in per cent: dots at 1.9, 2.4, 2.4, 2.9, 3.6, 3.8, 3.8 and 4.6, forming two clusters, with a dashed line at the median 3.25.

∴ across these eight quarters the annual CPI movement sat in two clusters — around 2–3% for 2024–25 and around 4% for 2025–26 — with a median of 3.25% and a range of 2.7%; prices rose faster in the later year than in the earlier one.

One last point: a dot plot shows the *distribution* and loses the time order. To see the rise from quarter to quarter, a time-series graph (the kind subtopic 3E reads) is the better display. Choosing the display to suit the question is part of planning.

Example 4 — unscaffolded

Nine students recorded how many text messages they received in one day: 4, 5, 5, 6, 6, 7, 7, 8, 46. Decide which measure of centre is more useful here, and report the data in two sentences.

The value 46 sits far from the other eight, so it is an **outlier** (Figure 8). With the outlier, the median is the 5th value, 6, and the mean is $94 \div 9 \approx 10.4$. Without it, the remaining eight values 4, 5, 5, 6, 6, 7, 7, 8 have median 6 and mean $48 \div 8 = 6$. The outlier leaves the median unchanged and drags the mean from 6 up to about 10.4 — a "typical" student did not receive 10 messages, so the mean misleads.

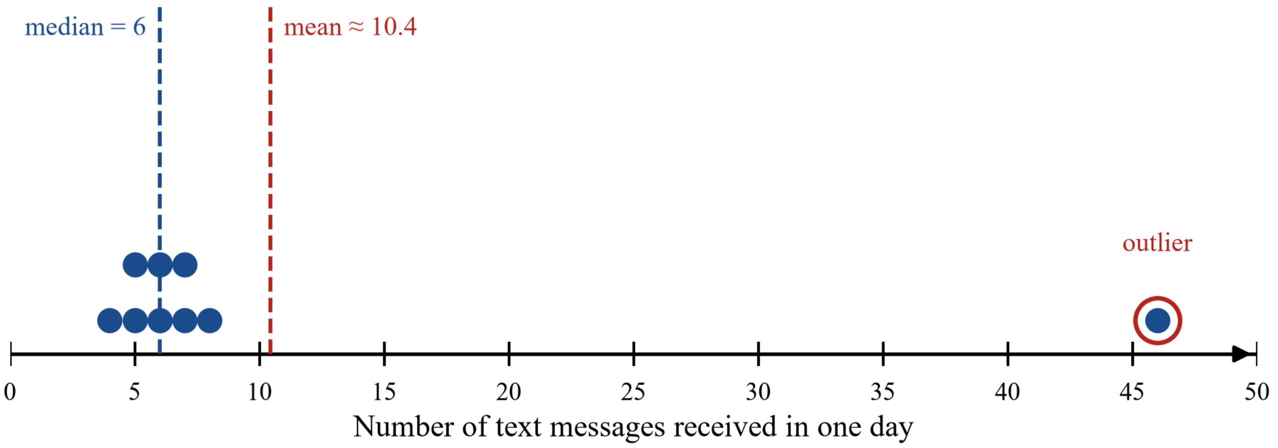

Figure 8. Dot plot of text messages received in one day: dots at 4, 5, 5, 6, 6, 7, 7 and 8, plus one circled outlier at 46; dashed lines show the median 6 and the mean about 10.4, which the outlier has pulled to the right of the data's cluster.

∴ the median (6) is the more useful centre, because the outlier of 46 pulls the mean up to about 10.4 while the median stays with the cluster; in words: "Most students received between 4 and 8 messages, with a median of 6; one student's 46 messages is an outlier that makes the mean a poor summary of this data."

Practice questions

Scaffolded

Q1. Say whether each question is a statistical question (answered with data that vary) or not. (a) How many pets does each student in our class have? (b) What is the capital of Victoria? (c) What are the reaction times of students in Year 7, measured with a ruler-drop test? (d) What day of the week is tomorrow? (e) On which day of the week do most students in our school walk to school? (f) How long is this pencil?

Q2. Name each variable as discrete or continuous. (a) the number of siblings a student has (b) the time taken to run 100 m (c) the number of goals kicked in a game (d) the height of a sunflower (e) the fuel in a car's tank, in litres (f) the number of eggs in a carton

Q3. Name each source of data as primary or secondary. (a) measuring the temperature of water in a beaker during an experiment (b) downloading rainfall records from the Bureau of Meteorology (c) asking every student in the class their favourite sport (d) reading population figures in an ABS publication (e) timing paper gliders your group built and flew (f) using figures from a newspaper article

Q4. These six stages of a statistical investigation are out of order. Write them in the correct order. analyse the data · collect data · organise and display the data · plan · pose a question · report findings

Q5. For each investigation, name the variable and say whether it is discrete or continuous. (a) counting the cars in the school car park at 8 am each day for a week (b) measuring how far a paper glider flies, in metres (c) measuring the height of each student in the class, in cm (d) counting the books in each student's school bag

Q6. The table gives practice data: the number of goals scored by a team in 17 games.

Goals	1	2	3	4	5
Number of games	2	5	6	3	1

(a) How many observations are in the data set?
(b) Find the mode.
(c) Find the median.
(d) Find the range.

Q7. Match each description to one of the words *symmetric*, *positive skew*, *negative skew*, *bi-modal*. (a) values cluster in the middle with even tails on both sides (b) a tail of values stretches toward the larger values (c) a tail of values stretches toward the smaller values (d) the values form two separate peaks

Q8. Look at the dot plot in Figure 6 (the ABS population data of Example 2). (a) How many states and territories are shown? (b) Which has the smallest population? (c) Which two are closest in population? (d) Read the median and the mean from the figure, in millions.

Unscaffolded

Q9. A group claims that a paper glider with rounded wings flies further than the same glider with pointed wings. Plan an investigation to test the claim. In your plan, state: the question; the variable and whether it is discrete or continuous; how many trials you will run; two things you will keep the same to make the test fair; and the display you will use.

Q10. Use the population data of Example 2. (a) Write the eight populations in order, smallest to largest. (b) Find the median population, in people. (c) Find the range, in people. (d) Find the mean population, in people. (e) How many of the eight states and territories have populations below the mean? (f) Is the distribution symmetric or skewed? If skewed, which way? Give one piece of evidence from the data.

Q11. The stem-and-leaf plot shows practice data: the times, in seconds, taken by 15 students to solve a puzzle.

```
1 | 2 5 8
2 | 1 2 4 4 7
3 | 1 3 5 8
```

```
4 | 1 4
5 | 2
```

Key: 1 | 2 means 12 seconds.

(a) How many students are in the data set?
(b) Find the median time.
(c) Find the range.
(d) Describe the shape of the distribution.
(e) Write one sentence reporting a finding in context.

Q12. Practice data from a ruler-drop test. Catch distances (cm) for five adults: 8, 9, 10, 10, 12. Catch distances (cm) for five Year 7 students: 5, 6, 6, 7, 8. (a) Find the median catch distance for each group. (b) Which group had the faster reactions? Give a reason. (c) Find the range for each group. (d) Write one sentence comparing the groups, using a centre and a spread.

Q13. Practice data: the number of books read this year by nine students is 2, 3, 3, 4, 4, 5, 5, 6, 32. (a) Find the median and the mean (to 1 decimal place). (b) Remove the value 32 and find the median and mean of the remaining eight values. (c) Is the mean in (a) a useful summary of this data? Explain which centre is more useful.

Q14. Use the CPI data of Example 3. (a) Find the median and the range of the eight values. (b) Name the two values that appear twice. What does the dot plot (Figure 7) show because of them? (c) Which quarter had the highest value, and what was it?

Q15. Mia plans an investigation of the question "Do teenagers have faster reaction times than adults?" Her plan: "I will use the ruler-drop test to measure the reaction times of 15 teenagers at my school." Explain why her plan cannot answer her question, and state one change that would fix it.

Q16. The issue is water use at home. (a) Write a statistical question about it. (b) Name the variable in your question and say whether it is discrete or continuous. (c) Choose a primary or a secondary source for the data, and name it.

Q17. A class survey asked "How many hours of homework do you do each week?" The results (n = 25) had median 4 hours, mean 5.2 hours, range 8 hours, and a positive skew. Write a two-sentence report of the findings, using at least two of these numbers and the shape.

Q18. The issue is the use of non-renewable resources (such as coal and oil) around the world. (a) Write a statistical question about it that secondary data could answer. (b) Name one place you would look for the data. (c) Name the variable and the display you would use.

Answers

Q1. (a) statistical; (b) not; (c) statistical; (d) not; (e) statistical; (f) not.

Q2. (a) discrete; (b) continuous; (c) discrete; (d) continuous; (e) continuous; (f) discrete.

Q3. (a) primary; (b) secondary; (c) primary; (d) secondary; (e) primary; (f) secondary.

Q4. Pose a question · plan · collect data · organise and display the data · analyse the data · report findings.

Q5. (a) number of cars, discrete; (b) flight distance in m, continuous; (c) height in cm, continuous; (d) number of books, discrete.

Q6. (a) 17; (b) 3 goals; (c) 3 goals; (d) 4 goals.

Q7. (a) symmetric; (b) positive skew; (c) negative skew; (d) bi-modal.

Q8. (a) 8; (b) the Northern Territory; (c) the ACT and Tasmania; (d) median ≈ 2.5 million, mean ≈ 3.5 million.

Q9. Example plan below.

Q10. (a) 267.5; 487.2; 579.1; 1 910.6; 3 076.5; 5 712.1; 7 121.9; 8 641.1 (thousands); (b) 2 493 550; (c) 8 373 600; (d) 3 474 500; (e) 5; (f) skewed — positively skewed.

Q11. (a) 15; (b) 27 s; (c) 40 s; (d) positive skew; (e) example sentence below.

Q12. (a) adults 10 cm, students 6 cm; (b) the students, shorter distances mean faster reactions; (c) adults 4 cm, students 3 cm; (d) example sentence below.

Q13. (a) median 4, mean ≈ 7.1; (b) median 4, mean 4; (c) no — the median is more useful.

Q14. (a) median 3.25, range 2.7; (b) 2.4 and 3.8 — two peaks, so the distribution is bi-modal; (c) Mar 2026, 4.6%.

Q15. She collects data from teenagers only, so there is nothing to compare them with; also measure adults the same way.

Q16. Example answers below.

Q17. Example report below.

Q18. Example answers below.

Fully-worked solutions

Q1. A statistical question must need data, and the data must vary. (a) Needs data that vary from student to student — statistical. (b) A fact looked up, no data — not statistical. (c) Needs reaction-time data that vary from person to person — statistical. (d) A fact about the calendar — not statistical. (e) Needs data that vary across the week and across students — statistical. (f) One pencil, one measurement, nothing varies — not statistical.

Q2. (a) Counted, whole numbers — discrete. (b) Measured, any value in a range — continuous. (c) Counted — discrete. (d) Measured — continuous. (e) Measured — continuous. (f) Counted — discrete.

Q3. (a) You measure it yourself — primary. (b) The Bureau of Meteorology collected it — secondary. (c) You ask the class yourself — primary. (d) The ABS collected it — secondary. (e) Your group flies and times the gliders — primary. (f) The newspaper collected and published it — secondary.

Q4. Figure 1 gives the order: pose a question, plan, collect data, organise and display the data, analyse the data, report findings.

Q5. (a) Variable: number of cars; counted — discrete. (b) Variable: flight distance in m; measured — continuous. (c) Variable: height in cm; measured — continuous. (d) Variable: number of books; counted — discrete.

Q6. (a) $2+5+6+3+1=17$ observations. (b) The largest count is 6 games, at 3 goals — mode 3. (c) The 9th value: 2 values are ≤ 1, 7 are ≤ 2, 13 are ≤ 3, so the 9th is 3 — median 3. (d) $5-1=4$ goals.

Q7. (a) symmetric; (b) positive skew (the tail points to the larger, positive, side); (c) negative skew; (d) bi-modal.

Q8. (a) Eight dots — 8 states and territories. (b) The dot nearest 0 is the Northern Territory, 0.27 million. (c) The ACT (0.49) and Tasmania (0.58) sit almost together, 0.09 apart — the closest pair. (d) Median ≈ 2.5 million (2.49), mean ≈ 3.5 million (3.47).

Q9. Example plan. Question: does a glider with rounded wings fly further than the same glider with pointed wings? Variable: flight distance in metres — continuous. Trials: 10 throws of each design, using the median of the 10. Fair test: the same thrower and the same throwing height for both designs, flown in the same room so there is no wind, measured with the same tape. Display: a dot plot for each design, side by side. The report will compare the two medians.

Q10. (a) In thousands: 267.5, 487.2, 579.1, 1 910.6, 3 076.5, 5 712.1, 7 121.9, 8 641.1. (b) The middle pair is 1 910.6 and 3 076.5: $(1910.6+3076.5)\div 2=2493.55$ thousand, so 2 493 550 people. (c) $8641.1-267.5=8373.6$ thousand, so 8 373 600 people. (d) The total is 27 796.0 thousand, and $27\,796.0\div 8=3474.5$ thousand, so 3 474 500 people. (e) The values below 3 474.5 thousand are NT, ACT, Tas, SA and WA, so 5 of the 8. (f) Positively skewed: five of the eight values sit below the mean, and the mean is well above the median — the tail runs toward the larger states.

Q11. (a) Count the leaves: $3+5+4+2+1=15$. (b) The 8th value is 27 s. (c) $52-12=40$ s. (d) Most times sit low, with a tail toward the larger times and one high value (52) — positive skew. (e) For example: "Most students solved the puzzle in 12 to 38 seconds, with a median of 27 seconds; one student took 52 seconds, well above the rest."

Q12. (a) Adults, ordered 8, 9, 10, 10, 12: median 10 cm. Students, ordered 5, 6, 6, 7, 8: median 6 cm. (b) The students: a shorter catch distance means a faster reaction, and 6 cm < 10 cm. (c) Adults: $12-8=4$ cm; students: $8-5=3$ cm. (d) For example: "The students reacted faster, with a median catch distance of 6 cm against the adults' 10 cm, and both groups were consistent, with ranges of 3 cm and 4 cm."

Q13. (a) Ordered, the 5th value is 4 — median 4. Total $=64$, so mean $=64\div 9\approx 7.1$. (b) Without 32: the middle pair of 4 and 4 gives median 4, and mean $=32\div 8=4$. (c) No. The one student who read 32 books drags the mean up to 7.1, above what seven of the nine actually read; the median of 4 stays with the cluster and is the more useful centre.

Q14. (a) Ordered: 1.9, 2.4, 2.4, 2.9, 3.6, 3.8, 3.8, 4.6. Median $=(2.9+3.6)\div 2=3.25$; range $=4.6-1.9=2.7$. (b) 2.4 and 3.8 each appear twice, giving the two peaks of a bi-modal distribution. (c) Mar 2026, at 4.6%.

Q15. Her question compares teenagers *with adults*, but her plan measures teenagers only, so there is no adult data to compare — the plan cannot answer the question. Fix: also measure a group of adults with the same ruler-drop test, the same number of trials each, and compare the two medians.

Q16. Example answers. (a) "How many litres of water does each household in our street use in a week?" (b) Variable: water used in litres — continuous (measured). (c) Secondary: the water company's usage records, or a published table; or primary: read each household's water meter on the same day for two weeks.

Q17. Example report: "A typical student did about 4 hours of homework each week (the median), and the hours varied widely, with a range of 8. The distribution had a positive skew, so the mean of 5.2 hours was pulled above the median by a few students who did many more hours."

Q18. Example answers. (a) "Which part of the world uses the most oil per person?" (b) A published energy report on the internet, or the ABS for Australian figures. (c) Variable: oil used per person (continuous); display: a dot plot of the values for each region, to compare shape, centre and spread.

Test

Chapter test T6 — Statistics

Assessed content descriptions	`VC2M7ST01` — subtopic 6A, Acquiring data; range, median, mean and mode; `VC2M7ST02` — subtopic 6B, Dot plots and stem-and-leaf plots: shape, centre and spread; `VC2M7ST03` — subtopic 6C, Planning and conducting a statistical investigation
Marks	**15** — Section A (6A): 5 marks · Section B (6B): 5 marks · Section C (6C): 5 marks
Time	15 minutes working time, plus 5 minutes reading time (20 minutes in all)
Equipment	Pen or pencil. No other equipment is needed for this paper.
Digital tools	Not needed and not permitted in this test. The content descriptions assessed here name no digital tool for data this small: a scientific calculator is not permitted, and every display is drawn by hand. All items are calculator-free.

> This is a school-designed paper. It is not a VCAA instrument. Achievement in the Victorian Curriculum is reported against the Level 7 achievement standard, not against a mark on this paper.

Answer every question. Show your working where a question asks for it. The marks for each part are shown in brackets.

Section	Subtopic	Content description	Marks	Questions
A	6A Acquiring data; range, median, mean and mode	`VC2M7ST01`	5	1–3
B	6B Dot plots and stem-and-leaf plots: shape, centre and spread	`VC2M7ST02`	5	4–5
C	6C Planning and conducting a statistical investigation	`VC2M7ST03`	5	6–7

Section A — Acquiring data; range, median, mean and mode (6A, `VC2M7ST01`)

Question 1 [3 marks]. One morning, a student counted the number of people in each of 11 cars arriving at the school gate:

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

a. Find the mode and the range. **[1 mark]**

b. Find the median. **[1 mark]**

c. Find the mean. **[1 mark]**

Question 2 [1 mark]. Players at a professional sports club are paid between $40 000 and $70 000 a year, but one star player is paid $2 000 000. The club wants one number to describe what a typical player is paid. Which measure of centre should it use?

- **A** the mean

- **B** the median
- **C** the mode
- **D** the range

Question 3 [1 mark]. Five students count the songs saved on their phones: 120, 140, 140, 150, 980. The median is 140 and the mean is 306; the value 980 is an outlier. Find the mean of the four values without the outlier, and say which centre the outlier changed more, the mean or the median.

Section B — Dot plots and stem-and-leaf plots: shape, centre and spread (6B, `VC2M7ST02`)

Question 4 [3 marks]. A class grew 12 sunflowers and measured each height (cm):

88, 95, 102, 104, 108, 110, 112, 115, 120, 124, 130, 135

a. Draw an ordered stem-and-leaf plot of the data, with a key. **[2 marks]**

b. Use your plot to find the median and the range. **[1 mark]**

Question 5 [2 marks].

a. The heights of another group of 20 students cluster between 130 and 160 cm, with one student at 188 cm. The median is 148 cm and the mean is 151 cm. Name the shape word — symmetric, positive skew, negative skew or bi-modal — that describes this distribution, and give one piece of evidence from the information given. **[1 mark]**

b. A Year 7 class and a Year 8 class both sit the same 20-question quiz. The teacher wants one display to compare the two sets of marks. Which display suits the data best?

- **A** a single dot plot of all the marks
- **B** a single ordered stem-and-leaf plot of all the marks
- **C** a back-to-back stem-and-leaf plot
- **D** a time-series graph

Section C — Planning and conducting a statistical investigation (6C, `VC2M7ST03`)

Section C asks you to analyse, interpret and report on data that has already been collected. Planning and conducting an investigation of your own is assessed in Investigation task I1, not in this test.

Question 6 [2 marks]. In one basketball match, the 9 players on a team scored these points:

8, 10, 11, 12, 12, 13, 14, 15, 60

a. Find the median and the mean (correct to 1 decimal place). **[1 mark]**

b. A reporter writes “a typical player scored about 17 points”. Explain which centre, the mean or the median, gives the truer picture of the team, and why. **[1 mark]**

Question 7 [3 marks]. A class surveyed 25 students, asking “How many minutes of exercise do you do each day?” The results have median 30 minutes, mean 38 minutes and range 55 minutes, and the distribution is positively skewed. Write a short report (2 or 3 sentences) that answers the question. Use at least two of the numbers and the shape.

End of test — 15 marks.

Solutions

Fully worked solutions for every question, followed by a compact answer key for self-checking.

Worked solutions

Question 1.

Put the values in order first: 1, 1, 2, 2, 2, 3, 3, 4, 4, 5, 6.

a. The value 2 appears three times, more than any other value, so the mode is 2 people. The range is the largest value minus the smallest:

$$6 - 1 = 5\,\text{people}$$

[1 mark]

b. There are 11 values, an odd number, so the median is the exact middle value — the 6th in the ordered list: 3 people. **[1 mark]**

c. Add every value, then divide by how many there are. The sum is 33, so

$$\text{mean} = 33 \div 11 = 3\,\text{people}$$

[1 mark]

Question 2.

The $2 000 000 payment is an outlier: it sits a long way above every other player's pay. The mean adds up every value, so the star player's pay drags the mean far above what a typical player receives; the median only looks at the middle of the ordered list and stays with the other players. The mode may not even exist (no pay rate need repeat), and the range is a measure of spread, not of centre. The answer is **B**, the median. **[1 mark]**

Question 3.

Without the outlier, the four remaining values are 120, 140, 140 and 150. Their sum is 550, so

$$\text{mean} = 550 \div 4 = 137.5$$

The outlier changed the **mean** more: the mean fell from 306 to 137.5, while the median stayed at 140 (the middle two of the four remaining values are both 140, so their mean is still 140). **[1 mark]**

Question 4.

a. The stems are the leading digits, 8 to 13; each leaf is the ones digit of one height. Written in order, with a key:

Stem	Leaf
8	8
9	5
10	2 4 8
11	0 2 5
12	0 4
13	0 5

Key: 8|8 means 88 cm.

Marks: 1 mark for the correct stems with all 12 leaves placed; 1 mark for the leaves ordered within each row and a correct key. **[2 marks]**

b. With 12 values the median is the mean of the 6th and 7th values. Reading down the leaves, those are 110 and 112:

$$(110 + 112) \div 2 = 111\,\text{cm}$$

The range reads straight off the first and last leaves:

$$135 - 88 = 47\,\text{cm}$$

[1 mark]

Question 5.

a. **Positive skew.** Either piece of evidence earns the mark: a tail of values stretches towards the larger heights (the one student at 188 cm), or the mean (151 cm) has been pulled above the median (148 cm) — both are what a positive skew looks like. **[1 mark]**

b. **C** — a back-to-back stem-and-leaf plot. It draws both classes against one shared stem column, so their centres and spreads can be compared directly. Options A and B mix the two classes into a single display, and a time-series graph is for change over time, not for comparing two groups. **[1 mark]**

Question 6.

a. The values in order are 8, 10, 11, 12, 12, 13, 14, 15, 60. With 9 values the median is the 5th: **12 points**. The sum is 155, so

$$\text{mean} = 155 \div 9 \approx 17.2\,\text{points (to 1 decimal place)}$$

[1 mark]

b. The 60-point player is an outlier — the next-highest score is only 15. That one score drags the mean up to 17.2, which is higher than every other player's score, so the mean describes almost nobody on the team. The median of 12 stays with the cluster of scores from 8 to 15, so the **median** gives the truer picture of a typical player. **[1 mark]**

Question 7.

Example report:

> A typical student exercised for about 30 minutes a day, which is the median. The times varied widely, with a range of 55 minutes, and the distribution was positively skewed: a few students who exercised for much longer pulled the mean (38 minutes) above the median.

Marks: 1 mark for a correct statement about the centre in context (a typical student, about 30 minutes, the median); 1 mark for a correct use of the spread or the shape (the range of 55 minutes shows the times varied, or the distribution was positively skewed); 1 mark for interpreting the difference between the mean and the median (the mean of 38 minutes is pulled above the median by the few students with much longer times). Accept any 2 or 3 sentences that use at least two of the given numbers and the shape correctly. **[3 marks]**

Answer key

Question	Answer
1a	Mode 2 people; range 5 people
1b	3 people
1c	3 people
2	B — the median
3	137.5; the mean changed more (the median stayed 140)
4a	Ordered stem-and-leaf plot as shown, key 8|8 means 88 cm
4b	Median 111 cm; range 47 cm
5a	Positive skew (tail towards the larger heights; mean above median)

Question	Answer
5b	C — a back-to-back stem-and-leaf plot
6a	Median 12 points; mean ≈ 17.2 points
6b	The median — the outlier of 60 drags the mean above almost every player's score
7	See worked solution (example report given)

Total: 15 marks — Section A (6A, `VC2M7ST01`): 5 · Section B (6B, `VC2M7ST02`): 5 · Section C (6C, `VC2M7ST03`): 5.