Cambridge IGCSE · Mathematics 0580 · Notes

# Averages and measures of spread

Read online at **pastopic.com/cambridge/mathematics-0580/notes/averages-and-spread**

Last checked 29 September 2026

Read first: **Fractions, decimals, ordering and the four operations**, **Equations and changing the subject**



## What you need to know

By the end of this topic you should be able to:

1. **Sort and tabulate data**: put results into a tally table or frequency table, and read and complete a two-way table.
2. **Read and interpret tables and statistical diagrams**, and draw sensible conclusions from them.
3. **Compare two sets of data** using a statistical measure: one average and one measure of spread (**Core:** the range; **Extended:** the range or the interquartile range).
4. **Know the limits of your data**: say when a conclusion goes too far, for example because the sample is small or the data is grouped.
5. **Work out the mean, median, mode and range** of data in a list, a stem-and-leaf diagram, a bar chart or a frequency table, and say which average suits a purpose best. **Core** data is never grouped.
6. **Extended only:** work out the **quartiles** and the **interquartile range** of a set of individual values.
7. **Extended only:** calculate an **estimate of the mean** of grouped data (discrete or continuous), using the midpoint of each class.
8. **Extended only:** name the **modal class** of a grouped frequency table.

Drawing stem-and-leaf diagrams, bar charts and pie charts is in the note on statistical charts and scatter diagrams. Reading the median and quartiles from a cumulative frequency graph, and histograms, are in the note on cumulative frequency and histograms.

## Understand it

### Kinds of data and tables

Data that are words or categories (favourite colour, type of pet) are **categorical**. Data that are numbers are **numerical**, and come in two kinds:

- **discrete** data are counted, so only certain values are possible (number of goals: $0, 1, 2, \ldots$);
- **continuous** data are measured, so any value in a range is possible (height, time, mass).

A **tally table** counts results as you go, in groups of five (|||| with a line through for 5). The count in each row is its **frequency**. A **two-way table** sorts people by two things at once. Every row and every column adds up to its total, which is how you fill in missing cells.

| | Walk | Bus | Car | Total |
|---|---|---|---|---|
| Girls | 6 | 5 | 3 | 14 |
| Boys | 4 | 9 | 3 | 16 |
| Total | 10 | 14 | 6 | 30 |

For example, if the "Boys, Bus" cell were blank, you would get it from $16 - 4 - 3 = 9$, or from $14 - 5 = 9$ down the column.

## The three averages and the range

An **average** is one number that represents the whole set. Take the five values $6, 2, 9, 4, 9$. First put them in order: $2, 4, 6, 9, 9$.

- **Mode**: the value that appears most often. Here it is $9$. A set can have two modes, or none if every value appears once.
- **Median**: the middle value **once the data are in order**. With $n$ values it is the $\frac{n+1}{2}$th value. Here $n = 5$, so it is the 3rd value, $6$. With an even number of values there are two middle values, and the median is halfway between them: for $3, 5, 8, 12$ it is $\frac{5+8}{2} = 6.5$.
- **Mean**: add all the values and divide by how many there are. Here $\frac{2+4+6+9+9}{5} = \frac{30}{5} = 6$.
- **Range**: largest value minus smallest value, $9 - 2 = 7$. It is a **measure of spread**: it says how spread out the data are, not where they sit. Always give it as one number.

Because $\text{mean} = \dfrac{\text{total}}{\text{number of values}}$, you can turn it round: $\text{total} = \text{mean} \times \text{number of values}$. This is the key to every question that gives you a mean and asks for a missing value.

## Which average to use

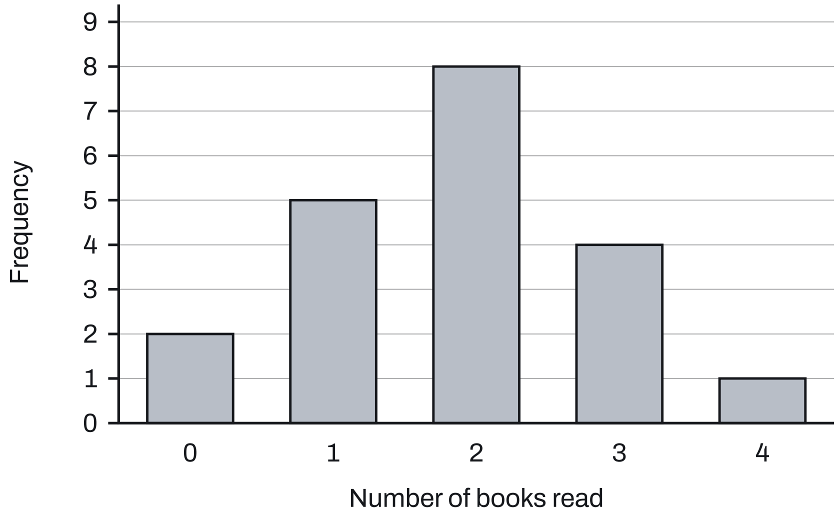
Each average has its own job:

- The **mode** is the only average for categorical data ("the most popular colour is blue"). It is also useful when you want the most common size, such as a shoe size to stock.
- The **median** is not pulled about by one very large or very small value (an **extreme value**, or outlier).
- The **mean** uses every value, so it is the fairest summary when there are no extreme values, but one extreme value drags it towards itself.

For example, five people earn these amounts, in thousands of dollars: $20, 22, 23, 25$ and $90$. The mean is $36$ thousand dollars, which is more than four of the five people earn, so it does not represent them well. The median, $23$ thousand dollars, is a better average here.

## Data in a frequency table or bar chart

A frequency table (or a bar chart) is a short way of writing a list with many repeats.




The row "2 books, frequency 8" means the value $2$ appears $8$ times in the list, so it adds $2 \times 8 = 16$ books to the total.

- **Mean**: multiply each value by its frequency, add these, and divide by the **total frequency**:
$$\text{mean} = \frac{\Sigma fx}{\Sigma f} = \frac{0 \times 2 + 1 \times 5 + 2 \times 8 + 3 \times 4 + 4 \times 1}{2+5+8+4+1} = \frac{37}{20} = 1.85 \text{ books.}$$
($\Sigma$ means "add up"; $x$ is a value and $f$ its frequency.)
- **Mode**: the value with the highest frequency, $2$ books. Not the frequency $8$.
- **Median**: $n = 20$, so the median is halfway between the 10th and 11th values. Count through the frequencies: values 1–2 are $0$, values 3–7 are $1$, values 8–15 are $2$. The 10th and 11th are both $2$, so the median is $2$ books.
- **Range**: $4 - 0 = 4$ books.

## Reading a stem-and-leaf diagram

In a stem-and-leaf diagram each value is split into a **stem** (here the tens digit) and a **leaf** (the units digit). The **key** tells you how to join them back together: "$3 \mid 0$ represents $30$ minutes". The leaves are in order, so the data are already sorted, and you can count along the rows to find any position.

**Homework time (minutes) of 11 students**

| | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | | |
| 2 | 3 | 5 | 8 | | lower quartile = 3rd value = 28 |
| | 4 | 5 | 6 | 7 | |
| 3 | 0 | 4 | 4 | 7 | median = 6th value = 34 |
| | 8 | 9 | 10 | | |
| 4 | 1 | 2 | 6 | | upper quartile = 9th value = 42 |
| | 11 | | | | |
| 5 | 5 | | | | |

Key: 3 | 0 represents 30 minutes

The values are $23, 25, 28, 30, 34, 34, 37, 41, 42, 46, 55$. The mode is $34$ minutes, the median is the 6th value, $34$ minutes, and the range is $55 - 23 = 32$ minutes. Always write the full value $(34)$, never just the leaf $(4)$.

## Quartiles and the interquartile range (Extended only)

**Extended only:** the median cuts the ordered data in half. The **quartiles** cut it into quarters:

- the **lower quartile** (LQ) is a quarter of the way through: the $\frac{n+1}{4}$th value;
- the **upper quartile** (UQ) is three quarters of the way through: the $\frac{3(n+1)}{4}$th value.

For the 11 homework times, $\frac{n+1}{4} = 3$ and $\frac{3(n+1)}{4} = 9$, so LQ $= 28$ and UQ $= 42$ (see the diagram above). The **interquartile range** is

$$\mathrm{IQR} = \mathrm{UQ} - \mathrm{LQ} = 42 - 28 = 14 \text{ minutes.}$$

The IQR is the range of the middle half of the data. Unlike the range, it ignores the extreme values at each end, so one unusual value does not change it much.

If $\frac{n+1}{4}$ is not a whole number, the quartile lies between two values. For $n = 10$ the lower quartile is the $2.75$th value, three quarters of the way from the 2nd to the 3rd. Mark schemes accept the nearby answers you get from the other common methods (such as taking the median of the lower half), so any sensible method earns the marks. Questions usually choose $n$ so that $n + 1$ is a multiple of $4$ (such as $11, 15$ or $19$), and then every method gives the same answer.

## Grouped data (Extended only)

**Extended only:** when data are grouped into classes, such as $10 < t \leqslant 20$, you no longer know the exact values. You cannot find the exact mean, but you can **estimate** it by pretending every value in a class sits at the class **midpoint**:

$$\text{midpoint} = \frac{\text{lower boundary} + \text{upper boundary}}{2}, \qquad \text{estimate of mean} = \frac{\Sigma f x}{\Sigma f} \text{ with } x = \text{midpoint}.$$

For example, $4$ values in $0 < x \leqslant 10$ and $6$ in $10 < x \leqslant 20$ have midpoints $5$ and $15$, so the estimate is $\frac{5\times4+15\times6}{10} = \frac{110}{10} = 11$. It is only an estimate because the real values in each class are unlikely to average exactly the midpoint.

- The **modal class** is the class with the highest frequency.
- The **class containing the median**: find the position of the middle value, then add up the frequencies until the running total reaches it.
- You cannot find the range exactly either: the largest and smallest values could be anywhere in their classes.

### Comparing two sets of data

**Core:** to compare two sets of data, make **two** comparisons:

1. **Average**: which set has the higher (or lower) mean or median, and what that means in the context;
2. **Spread**: which set has the larger range (**Extended:** or IQR), which means its values are **more varied**, less consistent.

For example: "Class A's marks were higher on average (median 64 compared with 58), and Class B's marks were more consistent (range 22 compared with 35)."

Be careful not to claim too much. A sample of 10 days may not represent the whole year; an estimate of the mean from grouped data is not exact; and a comparison tells you about the groups, not about any one person in them.

### Box-and-whisker plots (papers before 2025)

**Extended only:** papers up to 2023 sometimes showed a **box-and-whisker plot**, a diagram drawn from five numbers: the smallest value, LQ, median, UQ and largest value. The 2025–2027 syllabus leaves box-and-whisker plots out, so you won't meet them on your exam. If you practise an older question with one, the ideas are the same as above: the line inside the box is the median, the ends of the box are the quartiles, and the whiskers reach the smallest and largest values.

## Key facts and formulas

None of these is on the List of formulas printed in the exam paper, so you must know them all.

| Formula | Status |
|---|---|
| Mean $= \dfrac{\text{total of the values}}{\text{number of values}}$, so total $=$ mean $\times$ number of values | Learn it |
| Median: the $\dfrac{n+1}{2}$th value once the data are in order | Learn it |
| Mode: the most common value; range $=$ largest $-$ smallest | Learn it |
| Frequency table: mean $= \dfrac{\Sigma fx}{\Sigma f}$ | Learn it |
| Extended: lower quartile $= \dfrac{n+1}{4}$th value, upper quartile $= \dfrac{3(n+1)}{4}$th value | Learn it |
| Extended: interquartile range $=$ UQ $-$ LQ | Learn it |

| Formula | Status |
|---|---|
| Extended: grouped data, estimate of mean $= \dfrac{\Sigma fx}{\Sigma f}$ with $x$ the class midpoint | Learn it |
| Extended: midpoint $= \dfrac{\text{lower boundary} + \text{upper boundary}}{2}$ | Learn it |

## How to do it

In the Paper 1–4 questions we have tagged for this topic (91 questions, 2021–2025), the types came up this often. 26 of the 91 questions mix two or more types, so the counts add up to more than 91.

| Question type | Questions |
|---|---|
| Estimate the mean from a grouped frequency table (Extended) | 24 |
| Mode, median and range from a stem-and-leaf diagram | 21 |
| Mode, median, mean and range of a list | 19 |
| Averages from a frequency table, bar chart or pie chart | 17 |
| Find missing values from given averages and range | 10 |
| Compare, choose an average, or explain a change | 10 |
| Find an unknown from a given mean | 9 |
| Percentage of the data above or below a value | 5 |
| Quartiles and interquartile range (Extended) | 4 |
| Read a box-and-whisker plot (Extended, papers before 2025) | 4 |
| Class containing the median (Extended) | 2 |
| Range of consecutive integers (Extended) | 1 |

### 1. Estimate the mean from a grouped frequency table (24 questions)

**Extended only.** This is on nearly every Paper 4, usually next to a histogram or a cumulative frequency part, and is worth 4 marks.

1. **Write the midpoint of each class**: $10 < t \leqslant 20$ has midpoint $15$; $15 < t \leqslant 20$ has midpoint $17.5$. Keep the halves; don't round them.
2. **Multiply** each midpoint by its frequency.
3. **Add** these products to get $\Sigma fx$.
4. **Divide by the total frequency** $\Sigma f$ (the number of people or items), **not** by the number of classes.
5. Give the answer to 3 significant figures (such as $24.6$ for $24.583\ldots$), or exactly if the answer is exact.

**Variations you will meet:**

- **Unequal class widths** are common: work out each midpoint separately.
- **Frequencies read from a histogram or a cumulative frequency graph** first: frequency $=$ frequency density $\times$ class width; from a cumulative frequency graph, subtract neighbouring totals.
- **"Without calculating the new mean, find the change"** when one class's boundaries change: only that class's midpoint moves. The total changes by (change in midpoint) $\times$ (its frequency), so the mean changes by that amount $\div\Sigma f$ (Example 3(c)).

## 2. Mode, median and range from a stem-and-leaf diagram (21 questions)

1. **Read the key** to see what a stem and leaf stand for. A key such as $4 \mid 1$ representing $4.1$ m means the stem is the whole number part.
2. **Mode**: the leaf that repeats most in one row, written as the full value.
3. **Median**: count the leaves to get $n$, find the $\frac{n+1}{2}$th position, and count along the rows to it. For an even $n$, take the value halfway between the two middle ones.
4. **Range**: last leaf of the last row minus first leaf of the first row, as full values.

These are usually 1 mark each with no method marks, so check your count: the total number of leaves must equal $n$.

**Variations you will meet:**

- **Another value is added**: "Will the median increase or decrease?" If the new value is below the old median, the median goes down (or stays the same if there are repeated values at the middle); above it, the median goes up. Give the reason in words, such as "61 is greater than the median".

## 3. Mode, median, mean and range of a list (19 questions)

1. **Write the list in order** (smallest first) and count the values. Cross off each value as you copy it, so none is lost.
2. **Mode**: the most common value. **Median**: the middle value, or halfway between the two middle ones. **Range**: largest $-$ smallest.
3. **Mean**: add them all and divide by how many there are. Write the sum you divide, so a slip still earns the method mark.

**Variations you will meet:**

- **Change the units**: a range of heights in metres given in centimetres, such as $1.72 - 1.49 = 0.23$ m $= 23$ cm.
- **A mean of values you work out first** (for example, litres bought from the money spent and the price per litre): work out each value, then average them.

## 4. Averages from a frequency table, bar chart or pie chart (17 questions)

1. **Read the frequencies** (the heights of the bars, or each sector's angle as a fraction of $360^\circ$ times the total).

2. **Mean**: $\Sigma fx \div \Sigma f$. Adding a "$f \times x$" column to the table helps.
3. **Mode**: the **value** with the highest frequency.
4. **Median**: find the $\frac{n+1}{2}$th position, where $n = \Sigma f$, then add up the frequencies from the top until the running total reaches it.

**Variations you will meet:**

- **A median halfway between two values**: with $n = 30$ you need the 15th and 16th values. If the 15th is the last $1$ and the 16th is the first $2$, the median is $1.5$ (Quick check 2).
- **Categorical data** (such as favourite subjects): only the **mode** exists. The mean and median need numbers, and there is no range because categories have no largest and smallest.

## 5. Find missing values from given averages and range (10 questions)

1. **Draw boxes** for the values in order: $\square \leqslant \square \leqslant \square \ldots$
2. **Use the easy facts first**:
   - the **median** fixes the middle box (or the middle two);
   - the **mode** means that value appears at least twice, more than any other;
   - the **range** links the largest and smallest: largest $=$ smallest $+$ range;
   - "**no mode**" means every value is different.
3. **Use the mean last**: total $=$ mean $\times$ number of values, and the last box is the total minus the known values.
4. **Check** every condition with your final list.

**Variations you will meet:**

- **Two possible answers**: "the range of six numbers is 20, find the missing number" can give a new smallest value **or** a new largest value. Try both.
- **Missing leaves in a stem-and-leaf diagram**: the range gives the last leaf, the median gives the middle leaves, and the mean gives the one left.
- **Two unknowns with a known difference**: the total gives their sum; the sum and the difference together give each number.

## 6. Compare, choose an average, or explain a change (10 questions)

1. **Choose** as asked, then write **"because"** and name the statistic and its values: "Group A, because it has the higher mean (45 > 38)".
2. **Higher on average** $\rightarrow$ look at the mean or median. **More varied, less consistent** $\rightarrow$ look at the range or IQR: the larger one is more spread out.
3. **Why an average is unsuitable**: the mean is pulled by an extreme value (name the value); for categories only the mode can be found, because there are no numbers to add or put in order.

**Variations you will meet:**

- **"Will the median (or mean) increase or decrease?"** when a value is added: compare the new value with the old median (or mean).
- **A change in class width** that moves a midpoint: say which way the estimate moves and why.

## 7. Find an unknown from a given mean (9 questions)

1. **Total $=$ mean $\times$ number of values.**
2. **One more value added**: new value $=$ new total $-$ old total. For example, a mean of $12$ over $4$ games then a mean of $13$ over $5$ games: $65 - 48 = 17$.
3. **Unknown value or frequency in a table**: write $\Sigma fx = \text{mean} \times \Sigma f$ as an equation and solve it. If values $2$, $3$, $4$ have frequencies $3$, $n$, $2$ and the mean is $2.9$, then $6 + 3n + 8 = 2.9(5 + n)$, so $14 + 3n = 14.5 + 2.9n$, giving $0.1n = 0.5$ and $n = 5$. Check by working the mean out again: $\frac{29}{10} = 2.9$.

**Variations you will meet:**

- **An unknown $x$ that appears twice in a list**: the sum has $2x$ in it. Solve for $x$, then carry on with the complete list (for example to find quartiles).

## 8. Percentage of the data above or below a value (5 questions)

1. **Count** the values that fit. Watch the words: "less than 38" does not include $38$; "38 or more" does.
2. **Write the fraction** $\dfrac{\text{count}}{\text{total}}$ (this usually earns a mark on its own).
3. **Multiply by 100** to give the percentage.

## 9. Quartiles and interquartile range (4 questions)

**Extended only.**

1. **Order** the data (a stem-and-leaf diagram already is) and count $n$.
2. **LQ** $= \frac{n+1}{4}$th value, **UQ** $= \frac{3(n+1)}{4}$th value.
3. **IQR** $=$ UQ $-$ LQ.

When $\frac{n+1}{4}$ is not a whole number, take the value that far between (or halfway): mark schemes accept a small range of answers, as long as your method is sensible.

**Variations you will meet:**

- **"The lowest value in the top 75%"** (or similar percentages): that is the value at the lower quartile position.

### 10. Read a box-and-whisker plot (4 questions)

**Extended only, papers before 2025.** Box-and-whisker plots are not in the 2025–2027 syllabus. In older questions, read the median (the line in the box), the quartiles (the ends of the box) and the smallest and largest values (the ends of the whiskers); then range $=$ largest $-$ smallest and IQR $=$ UQ $-$ LQ, as above. A value higher than 75% of the data is above the upper quartile.

### 11. Class containing the median (2 questions)

**Extended only.**

1. **Find the median position**: for grouped data with $n$ items, use about the $\frac{n}{2}$th item (with $n = 60$, the 30th or 31st).
2. **Add the frequencies** from the first class until the running total reaches that position.
3. **Answer with the class**, such as $15 < t \leqslant 20$, not a single number. Don't assume it is the middle class of the table.

The **modal class** is simpler: the class with the highest frequency.

### 12. Range of consecutive integers (1 question)

**Extended only.** Write the list in algebra. $k$ consecutive integers starting at $a$ are $a, a+1, \ldots, a+k-1$, so the range is $(a+k-1) - a = k - 1$. Consecutive even (or odd) numbers go up in 2s, so their range is $2(k-1)$ (Quick check 5).

## Worked examples

The bold codes show where marks are earned, as in Cambridge mark schemes. The number is how many marks.

- **M1** method mark: for a correct method, even if the numbers slip.
- **A1** accuracy mark: for a correct answer; it needs the M mark before it.
- **B1** mark for a correct result on its own, with no method needed.

### Example 1

A hockey team scores these numbers of goals in 9 matches.

$3 \quad 0 \quad 2 \quad 5 \quad 1 \quad 2 \quad 4 \quad 2 \quad 6$

(a) Find the mode. **[1]**

(b) Find the median. **[2]**

(c) Calculate the mean. **[2]**

(d) Find the range. **[1]**

(e) The team plays a 10th match. The mean for all 10 matches is $2.9$ goals. Find the number of goals scored in the 10th match. **[2]**

In order: $0, 1, 2, 2, 2, 3, 4, 5, 6$.

**(a)** $2$ appears three times, so the mode is $2$. **B1**

**(b)** $n = 9$, so the median is the $\frac{9+1}{2} = 5$th value. **M1** for ordering the list. Median $= 2$. **A1**

**(c)** Total $= 0 + 1 + 2 + 2 + 2 + 3 + 4 + 5 + 6 = 25$. **M1** for the sum $\div 9$.

Mean $= \frac{25}{9} = 2.78$ (3 s.f.). **A1**

**(d)** $6 - 0 = 6$. **B1**

**(e)** New total $= 2.9 \times 10 = 29$. **M1** for $2.9 \times 10$ and the old total $25$.

10th match $= 29 - 25 = 4$ goals. **A1**

## Example 2

The stem-and-leaf diagram shows the time, in minutes, of 15 bus journeys on Route A.

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

Key: $2 \mid 5$ represents $25$ minutes

(a) Find the mode, the median and the range. **[3]**

(b) Work out the percentage of these journeys that took less than 20 minutes. **[2]**

(c) **Extended only:** work out the interquartile range. **[2]**

(d) On Route B, the median journey time is 24 minutes and the interquartile range is 7 minutes. Make two comparisons between the journey times on the two routes. **[2]**

**(a)** Mode $= 33$ minutes (three $3$s in the row with stem $3$). **B1**

$n = 15$, so the median is the 8th value. Counting: $14, 17, 18, 20, 22, 22, 25, 29$, so the median is $29$ minutes. **B1**

Range $= 48 - 14 = 34$ minutes. **B1**

**(b)** Journeys under 20 minutes: $14, 17, 18$, so $3$ out of $15$. **M1** for $\frac{3}{15}$.

$\frac{3}{15} \times 100 = 20\%$. **A1** ($20$ minutes itself is not "less than 20".)

**(c)** $\frac{n+1}{4} = 4$, so LQ $=$ 4th value $= 20$; $\frac{3(n+1)}{4} = 12$, so UQ $=$ 12th value $= 33$. **M1** for both quartiles.

IQR $= 33 - 20 = 13$ minutes. **A1**

**(d)** Journeys on Route A took longer on average, because its median $(29)$ is higher than Route B's $(24)$. **B1**

Journey times on Route A are more varied, because its IQR $(13)$ is larger than Route B's $(7)$. **B1**

## Example 3

**Extended only.** The table shows the times taken by 60 people to finish a puzzle.

| **Time ($t$ minutes)** | $0 < t \leqslant 10$ | $10 < t \leqslant 15$ | $15 < t \leqslant 20$ | $20 < t \leqslant 30$ | $30 < t \leqslant 50$ |
|---|---|---|---|---|---|
| Frequency | 8 | 14 | 22 | 12 | 4 |

(a) Calculate an estimate of the mean time. **[4]**

(b) Write down the class interval that contains the median. **[1]**

(c) Another puzzle club uses the same data, but its last class is $30 < t \leqslant 40$. Without calculating the new estimate, find how much smaller its estimate of the mean is. **[2]**

**(a)** Midpoints: $5, 12.5, 17.5, 25, 40$. **M1**

$\Sigma fx = 5 \times 8 + 12.5 \times 14 + 17.5 \times 22 + 25 \times 12 + 40 \times 4 = 40 + 175 + 385 + 300 + 160 = 1060$. **M1** for $\Sigma fx$ with each $x$ in its class.

Estimate $= \frac{1060}{60} = 17.7$ minutes (3 s.f.). **M1** (dependent on the previous M1) for $\div 60$, **A1**.

**(b)** The median is about the 30th or 31st person. Running totals: $8, 22, 44$, so both are in the third class: $15 < t \leqslant 20$. **B1**

**(c)** The last midpoint changes from $40$ to $35$, a drop of $5$ for each of the $4$ people, so $\Sigma fx$ drops by $20$. **M1**

The estimate drops by $\frac{20}{60} = 0.333$ minutes (3 s.f.). **A1**

## Example 4

Five positive integers have mode $7$, median $8$, mean $10$ and range $12$.

(a) Find the five integers. **[3]**

(b) A sixth integer is added. The mean of the six integers is $11$. Find the sixth integer and the new median. **[3]**

**(a)** In order: $\square\ \square\ 8\ \square\ \square$, as the median is the 3rd value. **B1**

The mode is $7$, and $7$ must appear more than any other value. Only the first two boxes are below $8$, so they are both $7$: $7, 7, 8, \square, \square$.

Range $12$: largest $= 7 + 12 = 19$. Total $= 10 \times 5 = 50$, so the 4th value is $50 - 7 - 7 - 8 - 19 = 9$. **M1** for using $10 \times 5$.

The integers are $7, 7, 8, 9, 19$. **A1** (Check: $9$ lies between $8$ and $19$, and $7$ is the only repeated value.)

**(b)** New total $= 11 \times 6 = 66$, so the sixth integer is $66 - 50 = 16$. **M1 A1**

In order: $7, 7, 8, 9, 16, 19$. The median is halfway between the 3rd and 4th values: $\frac{8+9}{2} = 8.5$. **B1**

## Common mistakes

- **Dividing by the wrong number.** For a frequency table the mean is $\Sigma fx \div \Sigma f$. Examiners often see $\Sigma fx$ divided by the number of classes (such as $5$), the mean of the frequencies, or the mean of the midpoints.
- **Using the wrong $x$ in grouped data.** Use the class **midpoints**, not the class widths or the boundaries, and keep midpoints such as $12.5$ exact.
- **Finding the median without ordering the list.** The middle of an unsorted list means nothing. Order first, and with an even count, take halfway between the two middle values.
- **Giving just the leaf.** From a stem-and-leaf diagram, the median is $34$, not $4$. Check the key, especially when it has decimals.
- **Giving the frequency instead of the value.** The mode of a frequency table is the value with the highest frequency, not that frequency. Likewise, the median of a table is a value, not the middle frequency or the middle row.
- **Mixing up the measures.** Examiners regularly see the mean given when the median was asked for, and the range given instead of the interquartile range. Read which one the question wants.
- **Writing the range as two numbers.** "$0$–$7$" is not a range; the range is the single number $7$.
- **Assuming the middle class contains the median.** Add up the frequencies to find it.
- **Losing a value.** When copying a list or drawing a stem-and-leaf diagram, count the values at the end to check none is missing.
- **Comparisons with no context or no statistic.** "Group A is bigger" earns nothing. Name the measure, its values, and what it means ("longer journeys on average").
- **Rounding too early.** Keep full values until the last line; a rounded midpoint or mean can push the answer outside the accepted range.

## How to get the marks

- **Grouped mean (4 marks):** one mark each for showing the midpoints, the $\Sigma fx$ calculation with the right values, and dividing by $\Sigma f$; then the answer. Write the full calculation, so a slip still leaves you three method marks.
- **One-mark answers** (mode, median, range from a diagram) have no method marks: only the exact answer scores, so check your counting.
- **"Calculate the mean"** with a list: write the sum and the division, which earns a method mark even if you mistype a value.
- **Missing values from a mean:** write "total $=$ mean $\times$ number" with the numbers in, such as $5 \times 13$ and $4 \times 12$. This is where the method marks are.
- **"You must show all your working"** means an answer with no working scores nothing, even if it is right.
- **Choose and explain** questions give one mark for the choice **and** a correct reason together: name the statistic ("higher median"), not just the choice.
- **"Write down"** means little or no working is needed; **"work out"** and **"calculate"** mean show the steps.
- **Accuracy:** give non-exact answers to 3 significant figures unless the question says otherwise, and give answers in the units asked for (minutes, cm).

## Quick check

1. Find the median and the mean of $12, 7, 9, 15, 7, 10$.
2. The table shows the number of pets owned by 30 students. Find the mode, the median and the mean.

| **Number of pets** | **0** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|---|
| Frequency | 6 | 9 | 7 | 3 | 5 |

3. The scores $1, 2$ and $3$ have frequencies $5, x$ and $3$. The mean score is $1.9$. Find $x$.
4. **Extended only:** the masses of 40 parcels are shown. Write down the modal class and calculate an estimate of the mean mass.

| **Mass ($m$ kg)** | $0 < m \leqslant 20$ | $20 < m \leqslant 40$ | $40 < m \leqslant 60$ | $60 < m \leqslant 80$ |
|---|---|---|---|---|
| Frequency | 5 | 11 | 16 | 8 |

5. **Extended only:** write down an expression for the range of $k$ consecutive odd numbers.

## Answers

1. In order: $7, 7, 9, 10, 12, 15$. Median $= \frac{9+10}{2} = 9.5$. Mean $= \frac{60}{6} = 10$.
2. Mode $= 1$ pet (frequency $9$). $n = 30$, so the median is halfway between the 15th and 16th values. Running totals $6, 15, 22$: the 15th value is $1$ and the 16th is $2$, so the median is $1.5$ pets. Mean $= \frac{0+9+14+9+20}{30} = \frac{52}{30} = 1.73$ pets (3 s.f.).
3. $\frac{1\times5+2x+3\times3}{5+x+3} = 1.9$, so $14 + 2x = 1.9(8 + x) = 15.2 + 1.9x$. Then $0.1x = 1.2$, so $x = 12$. (Check: $\frac{38}{20} = 1.9$.)
4. Modal class $40 < m \leqslant 60$. Midpoints $10, 30, 50, 70$: $\Sigma fx = 50 + 330 + 800 + 560 = 1740$, and $\frac{1740}{40} = 43.5$ kg.
5. The odd numbers go up in 2s: $a, a + 2, \ldots, a + 2(k - 1)$. Range $= 2(k - 1) = 2k - 2$.

## Past questions to try

- **0580/12/O/N/25 Q11**: range, mode, median and a percentage from a stem-and-leaf diagram, then the effect of a new value.
- **0580/32/F/M/25 Q10**: missing leaves from a given range, median and mean.
- **0580/22/F/M/25 Q16**: the interquartile range from a stem-and-leaf diagram.
- **0580/32/O/N/25 Q20**: comparing two cinemas using the mean and the range.