Averages and spread

S4 · Statistics · revision notes

Finding the mean, median, mode and range from a list or a frequency table, estimating the mean of grouped data, working backwards from a mean, and using an average and a measure of spread to compare two sets of data. This comes up on both tiers, often alongside a chart or table; Higher papers add harder problems that need algebra.

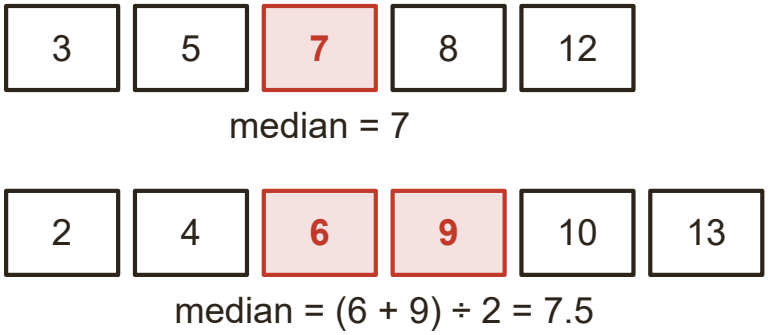
Key facts

- Mean = total of the values ÷ number of values
- Median = the $\frac{n+1}{2}$th value once the data are in order
- Mode = most common value; modal class = class with the highest frequency
- Range = largest value − smallest value
- Frequency table: mean = Σfx ÷ Σf
- Grouped data: estimated mean = Σ(f × midpoint) ÷ Σf
- Total = mean × number of values
- Compare with one average and one measure of spread, both in context

Notes

Mean, median, mode and range

- **Mode**: the most common value. There can be more than one mode, or none.
- **Median**: the middle value once the data are in order. For n values it is the $\frac{n+1}{2}$th value; with an even number of values it is halfway between the middle two.



Put the data in order first. With an even number of values, use the middle two.

- **Mean**: the total of the values ÷ the number of values.
- **Range**: largest value − smallest value. It measures how spread out the data are.
- An **outlier** is a value much bigger or smaller than the rest. It affects the mean and the range but hardly moves the median, so with an outlier the median is often the better average.
- Use the mode for non-numerical data, such as favourite colour.

Frequency tables

- Each value x has a frequency f: the number of times it occurs.
- Mean = Σfx ÷ Σf. Add an fx column, total it, then divide by the total frequency (not by the number of rows).
- The mode is the value with the highest frequency, not the frequency itself.
- For the median, find its position, $\frac{n+1}{2}$, then add up the frequencies from the top until you reach that position.

Grouped data

- In a grouped table you do not know the exact values, so the mean can only be **estimated**.
- Estimated mean = Σ(f × midpoint) ÷ Σf, using the midpoint of each class.
- The **modal class** is the class with the highest frequency. You can also find the class that contains the median.

Working with totals

- Total = mean × number of values. For a missing value, work out new total − old total.
- To combine two groups, add their totals and divide by the total number of values. Do not just average the two means unless the groups are the same size.
- If a frequency is unknown, write Σfx and Σf in terms of x, set Σfx ÷ Σf equal to the mean and solve the equation. Grade 7+ only

Comparing two sets of data

- Make two comparisons: one of an average (median or mean) and one of the spread (range).
- Say what each comparison means in context, e.g. 'On average, Class A scored higher' and 'Class B's scores were more consistent (smaller range)'.
- At Higher tier you may compare interquartile ranges instead of ranges (see Cumulative frequency and box plots). Grade 6+ only

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own, P = a process mark, C = a communication mark.

Find averages from a frequency table

3 to 4 marks **Grade 4**

1. Add an fx column: multiply each value by its frequency.
2. Add up the fx column and the frequency column.
3. Mean = Σfx ÷ Σf. Check that it lies between the smallest and largest values.
4. The mode is the value with the highest frequency.

Example. The table shows the number of pets owned by each of 30 students.

Number of pets	Frequency
0	7
1	10
2	8
3	4
4	1

(a) Write down the mode.

(b) Work out the mean number of pets.

Model answer.

(a) 1 pet (B1)

(b) Add an fx column (M1 for at least three correct products):

Pets (x)	Frequency (f)	fx
0	7	0
1	10	10
2	8	16
3	4	12
4	1	4
Total	30	42

42 ÷ 30 (M1)

= 1.4 pets (A1)

Work out an estimate for the mean of grouped data

4 to 5 marks | **Grade 5**

1. Find the midpoint of each class.
2. Multiply each midpoint by its frequency.
3. Add these up and divide by the total frequency.
4. Check that the answer lies inside the classes, and if asked, explain why it is only an estimate.

Example. The table shows information about the times, t minutes, that 40 people took to finish a puzzle.

Time (t minutes)	Frequency
$0 < t \leq 10$	6
$10 < t \leq 20$	14
$20 < t \leq 30$	12
$30 < t \leq 40$	8

(a) Work out an estimate for the mean time.
(b) Explain why your answer to (a) is only an estimate.

Model answer.

(a) Midpoints 5, 15, 25, 35 (M1)
6 × 5 + 14 × 15 + 12 × 25 + 8 × 35 = 30 + 210 + 300 + 280 = 820 (M1)
820 ÷ 40 (M1)
= 20.5 minutes (A1)
(b) The exact times are not known, so the midpoint of each class was used for every time in that class (C1)

Work out a combined mean or a missing value

2 to 3 marks | **Grade 5**

1. Turn each mean into a total: mean × number of values.
2. Add or subtract the totals, as the question needs.
3. Divide by the new number of values.

Example. In a test, the mean mark of 12 boys is 64 and the mean mark of 18 girls is 74.
Work out the mean mark of all 30 students.

Model answer.

12 × 64 = 768 or 18 × 74 = 1332 (M1)
(768 + 1332) ÷ 30 = 2100 ÷ 30 (M1)
= 70 (A1)
(64 + 74) ÷ 2 = 69 is wrong, because the two groups are different sizes.

Compare two distributions

2 marks **Grade 4**

1. Compare the averages: say which is higher and what that means in context.
2. Compare the spreads: the smaller range means the data are more consistent.
3. Write each comparison as a full sentence that uses the numbers.

Example. Two basketball teams each played 10 matches. The table shows information about the points they scored.

	Median	Range
Team A	24 points	15 points
Team B	19 points	6 points

Compare the numbers of points scored by the two teams.

Model answer.

On average Team A scored more points, because its median (24) is higher than Team B's (19) (C1)

Team B's scores were more consistent, because its range (6) is smaller than Team A's (15) (C1)

Find an unknown frequency from the mean 4 marks **Grade 7**

1. Write the total frequency, Σf, in terms of x.
2. Write Σfx in terms of x.
3. Set Σfx ÷ Σf equal to the given mean and multiply out.
4. Solve the linear equation and check your answer.

Example. The table shows the scores of some players in a game.

Score	Frequency
1	5
2	x
3	7
4	3

The mean score is 2.4
Work out the value of x.

Model answer.

Σf = 15 + x and Σfx = 5 + 2x + 21 + 12 = 38 + 2x (P1)

$$\frac{38 + 2x}{15 + x} = 2.4$$ (P1)

38 + 2x = 36 + 2.4x, so 2 = 0.4x (P1)

x = 5 (A1)

Check: Σfx = 48 and Σf = 20, and 48 ÷ 20 = 2.4

Don’t lose marks

- Forgetting to put the data in order before finding the median.
- Giving the position $\frac{n+1}{2}$ as the median, instead of the value in that position.
- Giving the highest frequency as the mode instead of the value that has that frequency.
- Dividing Σfx by the number of rows in the table instead of by the total frequency.
- Using the class width or one end of a class instead of the midpoint when estimating a mean.
- Averaging two means when the groups are different sizes.

More tips

Memory tricks

- Mode = most, median = middle, mean = share it out equally, range = spread.
- Sense check: a mean must lie between the smallest and largest values. For grouped data it must lie inside the classes that contain data.
- Mean × number = total. Most 'missing value' problems start here.
- On a calculator paper you can check a frequency table mean with your calculator's statistics mode, but still write out Σfx and Σf for the method marks.
- Median of 30 values: the $\frac{30+1}{2}$ = 15.5th value, which is halfway between the 15th and 16th values.

Exam technique

- Show your fx column or list of products. Method marks can still be given if you slip later.
- In 'compare' questions, give two comparisons (an average and the spread), each with the numbers and in the context of the question.
- Use 'on average' when comparing averages and 'more consistent' or 'less spread out' when comparing ranges.
- When asked which average is best, name it and give a reason linked to the data, e.g. 'the median, because 95 is an outlier that makes the mean too high'.
- Do not round in the middle of a calculation. Round only the final answer, if the question asks you to.

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade 1 — **Find the mode and range of a list**
The mode is the most common value; the range is the largest value minus the smallest value.

Grade 2 — **Find the median of a list**
Put the values in order and find the middle one, or halfway between the middle two.

Grade 3 — **Calculate the mean of a list**
Add up all the values and divide by how many values there are.

Grade 4 — **Find averages from a frequency table**
Mean = total of the fx column ÷ total frequency; the mode is the value with the highest frequency.

Grade 4 — **Compare two data sets using average and spread**
Compare an average and the range, and say what each comparison means in context.

Grade 5 — **Estimate the mean of grouped data**
Use the midpoint of each class as the value of every item in that class.

Grade 5 — **Solve problems with totals and combined means**
Use total = mean × number of values to find a missing value or the mean of two groups together.

Grade 7 — **Find an unknown frequency from a mean**
Write the mean as an algebraic fraction, set it equal to the given mean and solve the equation.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.