Statistics & Probability · Unit S06 · Foundations

The Three Averages

An average is one number standing in for a whole list. There are three of them, they usually disagree, and choosing between them is most of the skill. This unit works all three out from a raw list, from a frequency table and from grouped classes — and then watches what one extreme value does to each.

Foundations (MYP 1–3) · about 50 minutes · 60 practice questions · Criteria A, B, C and D

BY THE END OF THIS UNIT YOU CAN

• **Calculate** the **mean**, the **median** and the **mode** of a raw list of values

• **Use** the median position **(n + 1) ÷ 2** for an odd and for an even number of values

• **Calculate** a mean straight from a frequency table as **Σfx ÷ Σf**, without re-listing the data

• **State** the **modal class** of grouped data and **estimate** its mean from the class midpoints

• **Describe** why a mean from grouped data is an estimate and never the exact answer

• **Recompute** the mean and the median after one extreme value joins, and **describe** which one moved

• **Choose** the right average for a situation, and **describe** the reason

• **Round** an average sensibly and **state** its unit

THE ONE IDEA BEHIND ALL OF IT

Three averages, three different questions. The **mean** asks “if everyone had the same, how much would each get?” The **median** asks “who is standing in the middle?” The **mode** asks “which answer came up most often?” None of them is the average; each is *an* average. So the useful question is never “what is the average?” but **“which average, and why that one?”**

$$\text{mean} = \Sigma fx \div \Sigma f$$

and every average must land between the smallest and the largest value — that is the check

WHERE THIS UNIT SITS

What this unit assumes. Adding a list and dividing by how many there are (Number N02 and N03 — fraction and decimal arithmetic), and rounding an answer sensibly to 3 significant figures (Number N04 *Round It Right*). Neither is retaught here; both are used on every page.

What it builds on in this strand. S04 *Picture the Data* taught you to read values out of a chart. Here you average the values you read. The frequency tables are exactly the tables S01 built.

What comes next. S07 *How Spread Out?* adds the second half of every comparison: an average on its own says where the data sits, but not how tightly it is packed. S17 returns to grouped data when the values are measured rather than counted.

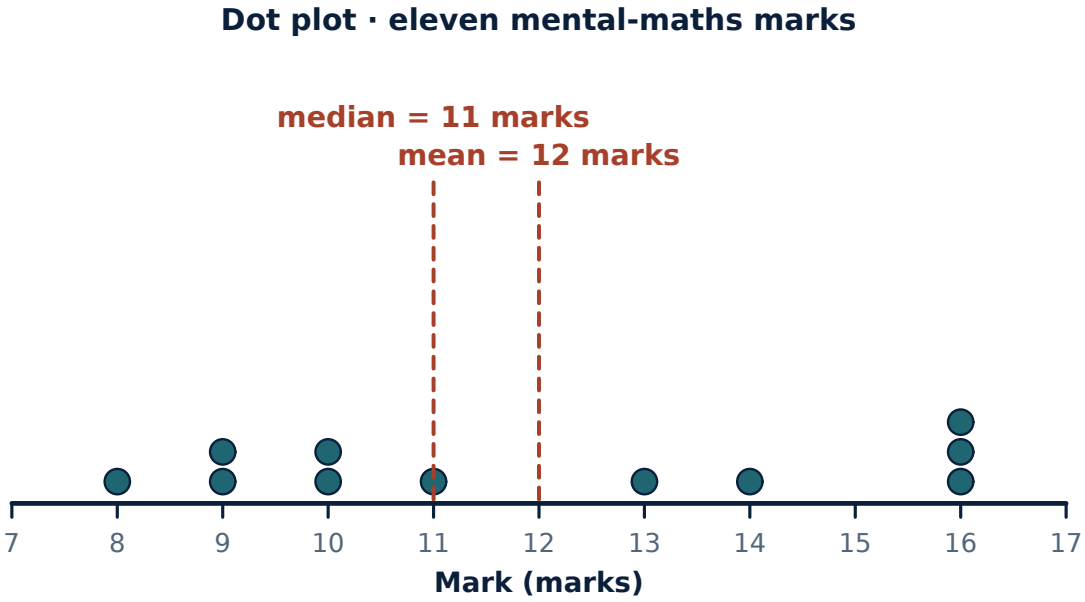
Where the numbers in this unit come from. Eleven marks, printed in full; a frequency table of 30 learners, printed in full; a grouped table of 30 players, printed in full; nine sponsored-walk amounts, printed in full. Every average on this page can be checked by hand against the data it came from, and that is deliberate.

CONCEPT 1 · THREE AVERAGES, ONE LIST

mean, median and mode

The **mean** is the total of all the values divided by how many values there are — the “share it out equally” average. The **median** is the middle value once the list is in order. The **mode** is the value that occurs most often. All three are called averages, all three are single numbers standing in for the whole list, and on almost any real dataset all three are different. Write down which one you are finding before you start calculating, because the three take three different routes.

> **EXAMPLE · The raw list, in full.** Eleven learners sat a mental-maths quiz marked out of 20. Their marks, in the order the teacher read them out, were: 16, 8, 10, 11, 9, 16, 10, 9, 13, 16, 14.



Dashed: the mean and the median. Tallest stack: the mode.

Every dot is one learner. The tallest stack is the mode; the two dashed lines are the mean and the median, and they are not in the same place.

Worked example 1 · Three answers, one list

1	Write the eleven marks in order, smallest first.	**8, 9, 9, 10, 10, 11, 13, 14, 16, 16, 16**
2	Add them all up.	**total = 132**
3	Divide the total by how many values there are.	**mean = 132 ÷ 11 = 12 marks**
4	Find the median position: (n + 1) ÷ 2 with n = 11.	**position 6**
5	Count in to the 6th value of the ordered list.	**median = 11 marks**
6	Look for the value that occurs most often.	**mode = 16 marks**
7	Check all three lie between the smallest and the largest mark.	**8 to 16**

Three averages, one list, three different numbers — and not one of them is wrong. Set the working out in this order every time: order the list, then mean, then median, then mode, then the check that all three sit inside the data.

EXAMPLE. The eleven marks in the dot plot above give three different answers: the mean is **12**, the median is **11** and the mode is **16**. Same eleven learners, same eleven marks, three honest answers to three different questions.

CONCEPT 2 · FINDING THE MIDDLE

order first, then count in

The median needs the list **in order** — smallest to largest — before anything else happens. Then the middle value sits at position **(n + 1) ÷ 2**. With *n* odd that lands on a value: for 11 values it is the 6th. With *n* even it lands halfway between two positions: for 10 values it is position 5.5, so the median is the mean of the 5th and 6th values, and it need not be a value in the list at all. The mode needs no ordering, but it can be shared by two values, and if every value occurs once there is no mode.

EXAMPLE. For the eleven marks, (11 + 1) ÷ 2 = 6, so the median is the **6th value of the ordered list**, which is 11. For the ten sponsored-walk amounts later in this unit, (10 + 1) ÷ 2 = 5.5, so the median is halfway between the 5th and the 6th values.

CONCEPT 3 · AVERAGES FROM A TABLE

Σfx ÷ Σf

A frequency table is a list with the repeats collected up, so you never have to re-write it out. Add a third column, **f × x**: each value multiplied by how many times it occurred. Then **mean = Σfx ÷ Σf** — the total of the f × x column divided by the total of the frequency column. The symbol Σ just means “add up”. The **mode** is the value with the largest frequency, not the frequency itself, and the **median** is found by counting down the frequency column until you reach the middle position.

Frequency table · a five-word spelling test, 30 learners

Correct answers, x	Frequency, f	f × x
0	2	0
1	6	6
2	10	20
3	6	18
4	5	20
5	1	5
Total	30	69

Divide the f × x total by the frequency total.

The f × x column does the repeating for you, so the 30 scores are never written out again.

Complete the f × x column

Correct answers, x	Frequency, f	f × x
0	2	0
1	6	6
2	10	
3	6	18
4	5	
5	1	5
Total	30	

Three boxes to fill: two products and the total.

Work each row out from the table itself. The total is the check, not the source.

Worked example 2 · From the table

1	Add an f × x column: each score multiplied by its frequency.	**six rows**
2	Row x = 0: 0 × 2.	**f × x = 0**

3	Row x = 2: 2 × 10.	**f × x = 20**
4	Row x = 5: 5 × 1.	**f × x = 5**
5	Add the f × x column.	**Σfx = 69**
6	Add the frequency column.	**Σf = 30**
7	Divide: mean = Σfx ÷ Σf.	**69 ÷ 30 = 2.3 correct answers**

The whole point is that the 30 scores are never written out again. The f × x column does the repeating for you, and the divisor is Σf = 30 — the number of learners — never the number of rows.

EXAMPLE. From the spelling table: Σfx = **69** and Σf = **30**, so the mean is 69 ÷ 30 = **2.3** correct answers. Notice that the mean is not a possible score — no learner got 2.3 answers right — and that is perfectly normal.

CONCEPT 4 · WHEN THE DATA IS GROUPED

midpoints, and the word estimate

Grouped data has been sorted into classes and the raw values thrown away. You can still say which class is fullest — the **modal class** — and you can still get a mean, but only by **assuming** every value in a class sits at that class's **midpoint**. The estimated mean is **Σfm ÷ Σf**, where *m* is the midpoint of each class. It is an **estimate**, and the word is part of the answer: the real values are gone, so the real mean cannot be recovered from this table.

Grouped table · goals scored by 30 players in a season

Goals scored, x	Midpoint	Frequency, f
0 ≤ x < 5	2.5	2
5 ≤ x < 10	7.5	6
10 ≤ x < 15	12.5	11
15 ≤ x < 20	17.5	8
20 ≤ x < 25	22.5	3
Total		30

The raw counts are gone, so a mean can only be estimated.

Five classes, five midpoints. The estimated mean is the f × m total divided by 30.

Worked example 3 · Grouped goals

1	Find the midpoint of each class: halfway between its two boundaries.	**2.5, 7.5, 12.5, 17.5, 22.5**
2	Class 0 ≤ x < 5: midpoint 2.5, frequency 2, so f × m.	**f × m = 5**
3	Class 10 ≤ x < 15: midpoint 12.5, frequency 11, so f × m.	**f × m = 137.5**
4	Add the f × m column.	**Σfm = 395**
5	Divide by the total frequency.	**395 ÷ 30 = 13.2 goals (3 s.f.)**
6	Name the fullest class.	**modal class 10 ≤ x < 15**
7	Say why the mean is only an estimate.	**the raw goal counts are gone**

*Every number in that table is a class, not a value, so the mean it gives is an **estimate**. Say the word in the answer. The class 0 ≤ x < 5 is written with boundaries, so its midpoint is 2.5 — and since goals are whole numbers, that midpoint is itself an assumption, which is one more reason the answer is an estimate.*

EXAMPLE. From the grouped table: Σfm = **395** and Σf = **30**, so the estimated mean is 395 ÷ 30 = **13.2** goals to 3 significant figures. The modal class is **10 ≤ x < 15**. Both answers carry the word “estimated” or the word “class” for a reason.

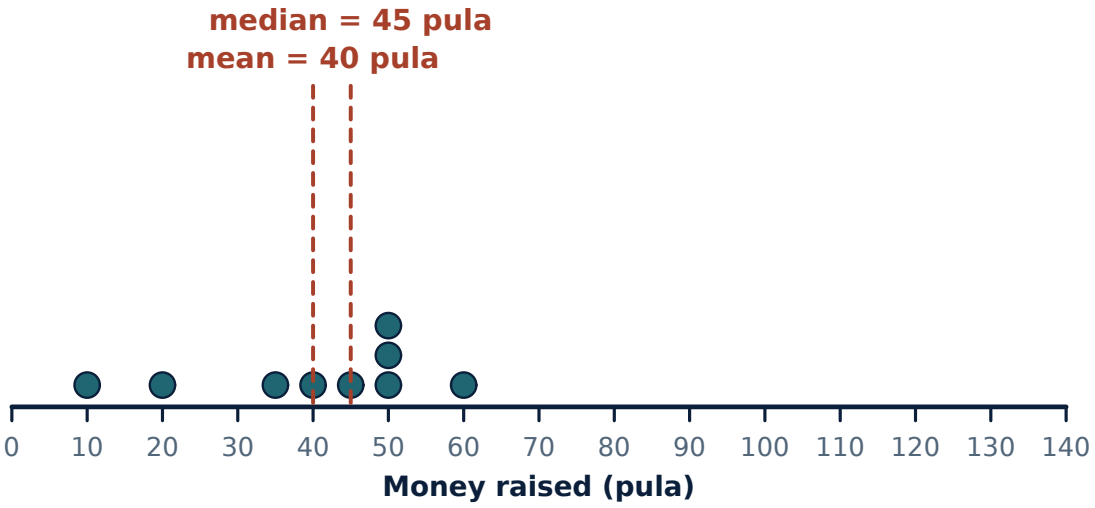
CONCEPT 5 · ONE EXTREME VALUE

what an outlier does to each average

One value far away from the rest is an **extreme value**, or **outlier**. The mean is built from the total, so every value pulls on it and a far-away value pulls hard. The median is built from a *position*, so a far-away value only shifts it along the order by one place. That is the whole difference: an outlier moves the mean a long way and the median hardly at all. Neither answer is wrong — but only one of them still describes a typical value.

EXAMPLE · The raw list, in full. Nine learners walked for the school fund. The money each of them raised, in pula, was: 50, 20, 40, 60, 45, 50, 50, 35, 10. A tenth learner then hands in a card for 130 pula.

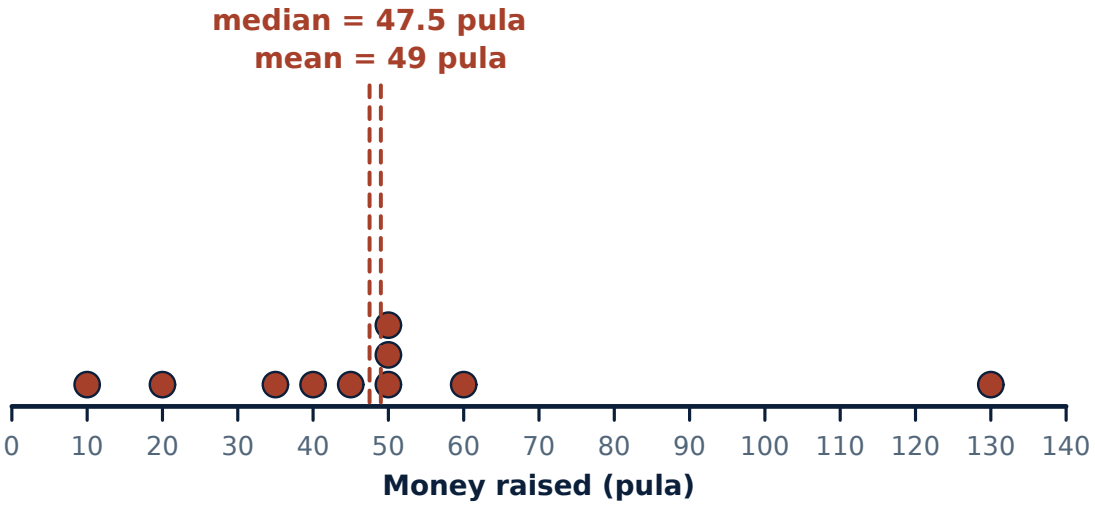
Before · nine sponsored-walk amounts



Nine amounts, before the tenth card arrives.

The two markers are already apart, because the amounts are not evenly spread.

After · the same nine, plus one amount of 130 pula



The same nine, plus one amount of 130 pula.

Only one of the two markers really moves — and it moves past most of the data.

Worked example 4 · The visitor

1	Write the nine amounts in order.	**10, 20, 35, 40, 45, 50, 50, 50, 60**
2	Mean before: total ÷ 9.	**360 ÷ 9 = 40 pula**
3	Median before: (9 + 1) ÷ 2 = position 5.	**45 pula**
4	Add the tenth amount of 130 pula and re-order.	**10, 20, 35, 40, 45, 50, 50, 50, 60, 130**
5	Mean after: total ÷ 10.	**490 ÷ 10 = 49 pula**
6	Median after: (10 + 1) ÷ 2 = position 5.5, so halfway between the 5th and 6th values.	**47.5 pula**
7	Describe which average moved, and by how much.	**mean by 9, median by 2.5**

One value moved the mean 9 pula and the median 2.5 pula. The mean is built from the total, so it feels every pula of the newcomer; the median is built from a position, so it only slides along the order.

EXAMPLE. Nine amounts have mean **40** pula and median **45** pula. One more amount of 130 pula arrives, and the mean jumps to **49** while the median moves only to **47.5**. The mean moved 9 pula; the median moved 2.5.

CONCEPT 6 · WHICH AVERAGE, AND WHY

choosing, and saying why

Mode for words and categories, and for anything where “most popular” is the question — a shop orders the modal shoe size, not the mean one. **Median** when one or two extreme values would drag the mean somewhere no one actually is — pay, prices, house sizes. **Mean** when the values are reasonably even and you want every one of them to count. Say which one you chose and why: an average quoted without its name is an invitation to be misled, which is where S03 came in.

Worked example 5 · Which average?

1	A shop must decide which shoe sizes to order for next term.	**mode — the size that sells most**

2	Give the reason.	**the mean size might not be a size the shop can even buy**
3	A small workshop is asked what a worker there typically earns, and one person earns many times what the others do.	**median — it ignores the extreme value**
4	Give the reason.	**the mean is dragged up by the one large pay**
5	A teacher wants one number for a class test where the marks are spread evenly with no extreme values.	**mean — every mark counts**
6	Give the reason.	**no value is far enough out to distort the total**

Choosing is the assessed skill, and the reason earns as much as the choice. Categories take the mode; extreme values take the median; even, complete data takes the mean.

EXAMPLE. Which average would you quote for the ten walk amounts? The mean is 49, the median is 47.5 and the mode is 50. If you want to describe what a typical learner raised, the mean is the one number of the three that no learner is anywhere near — one big sponsorship has moved it.

THE AVERAGING ROUTINE

1. Which average is asked for? Mean, median or mode. Write the word down first: the three take three different routes and mixing them up costs the whole mark.

2. Put the list in order. Only the median strictly needs it, but ordering also makes the mode obvious and makes a silly mean easy to spot.

3. Do the arithmetic that average asks for. Mean: total ÷ how many. From a table: Σfx ÷ Σf. Grouped: Σfm ÷ Σf, and call it an estimate. Median: position (n + 1) ÷ 2. Mode: biggest frequency, then read off the value.

4. Check it sits inside the data, and give it its unit. An average must lie between the smallest and the largest value. If it does not, the arithmetic is wrong. Then say what it is: 12 marks, 2.3 correct answers, 49 pula.

THE MISTAKES THAT COST THE MOST MARKS

Trap 1 · Dividing Σfx by the number of rows instead of Σf

Wrong. “Σfx = 69 and the table has 6 rows, so the mean is 69 ÷ 6 = 11.5.”

Right. The divisor is **Σf = 30**, the number of *learners*: 69 ÷ 30 = **2.3**.

A frequency table has one row per *value* and one column of *how many*. The number of rows is a fact about the table; the mean is a fact about the learners. The check that catches it every time: an average must lie between the smallest and the largest value, and 11.5 is not even a possible score.

Trap 2 · Taking the middle of the list as it was written down

Wrong. “For 12, 5, 20, 9, 7 the middle one is 20, so the median is 20.”

Right. Order the list first: 5, 7, 9, 12, 20. The middle value is now **9**, so the median is 9.

The median is defined on the *ordered* list, and a raw list almost never arrives in order. Ordering costs ten seconds and is the only thing standing between you and a wrong median. Write the ordered list out on the answer line — it is also evidence of method.

Trap 3 · Reporting the frequency instead of the value as the mode

Wrong. “The biggest number in the spelling table is 10, so the mode is 10.”

Right. 10 is the *frequency*. The mode is the **value** that goes with it: 10 learners scored **2**, so the mode is 2 correct answers.

Say the sentence out loud: “the most common *score* was …”. The mode always answers in the units of the variable — correct answers, shoe sizes, goals — never in learners. The same slip turns up in bar charts, where the tallest bar is read as the answer instead of the label underneath it.

PRACTICE · 60 QUESTIONS

Answers are at the back. Work each one out before you look.

Concept 1 · Three averages, one list

P1 (Fluency)	**Calculate** the mean of 5, 8, 11, 12, 14.
P2 (Fluency)	**State** the mode of 3, 7, 7, 9, 12.
P3 (Fluency)	The list 4, 6, 9, 11, 20 is already in order. **State** the median.
P4 (Fluency)	The eleven marks are printed above. **State** the mode.
P5 (Applied)	**Calculate** the mean of the eleven marks, and **state** the total you divided.
P6 (Applied)	**State** the median of the eleven marks and **describe** how you found it.
P7 (Applied)	**Calculate** the mean of 12, 15, 18, 21 and **describe** why an average need not be one of the values.
P8 (Applied)	A list of 6 values has a mean of 7. **Calculate** the total of the six values.
P9 (Challenge)	Five values have a mean of 9. Four of them are 6, 8, 10 and 12. **Calculate** the fifth value.
P10 (Challenge)	**Describe** the difference between what the mean uses and what the median uses.

Concept 2 · Finding the middle

P11 (Fluency)	**State** the median of 12, 9, 15.
P12 (Fluency)	**Calculate** the median of 4, 7, 9, 10.
P13 (Fluency)	A list has 15 values in order. **State** the position of the median.

P14 (Fluency)	**State** the mode of 2, 3, 3, 5, 7, 7.
P15 (Applied)	The nine sponsored-walk amounts are printed below. **Calculate** their median.
P16 (Applied)	A list has 12 values. **Describe** where its median lies.
P17 (Applied)	**Order** the list 14, 8, 22, 11, 9, 30, 17 and **state** the median.
P18 (Applied)	One more amount of 130 pula joins the nine. **Calculate** the median of the ten amounts.
P19 (Challenge)	A learner reads the median of 14, 8, 22, 11, 9, 30, 17 straight off as 11. **Describe** the error and **state** the correct median.
P20 (Challenge)	A set of 8 values has a median of 12, and the two middle values are equal. **State** those two values and **describe** why.

Concept 3 · Averages from a table

P21 (Fluency)	The spelling table is printed above. **State** the frequency of the score 2.
P22 (Fluency)	The spelling table is printed above. **Calculate** f × x for the row x = 4.
P23 (Fluency)	The spelling table is printed above. **State** the mode.
P24 (Fluency)	The spelling table is printed above. **State** Σf.
P25 (Applied)	**Calculate** Σfx for the spelling table.
P26 (Applied)	**Calculate** the mean score of the 30 learners.
P27 (Applied)	**Calculate** the median score of the 30 learners from the table.
P28 (Applied)	**Describe** why the mode of the spelling table is 2 and not 10.

P29 (Challenge)	A learner computes the mean of the spelling table as 69 ÷ 6 = 11.5. **Describe** the error and **state** the correct mean.
P30 (Challenge)	**Calculate** how many of the 30 learners scored 3 or more, and **state** which rows you added.

Concept 4 · When the data is grouped

P31 (Fluency)	The grouped goals table is printed above. **State** the modal class.
P32 (Fluency)	**State** the midpoint of the class $5 \leq x < 10$.
P33 (Fluency)	**State** how many players are in the class $15 \leq x < 20$.
P34 (Fluency)	The grouped goals table is printed above. **State** Σf.
P35 (Applied)	**Calculate** $f \times m$ for the class $10 \leq x < 15$.
P36 (Applied)	**Calculate** Σfm for the grouped goals table.
P37 (Applied)	**Calculate** the estimated mean number of goals, to 3 significant figures.
P38 (Applied)	**Describe** why the mean from the grouped table is only an estimate.
P39 (Challenge)	**Determine** which class holds the median of the 30 players, and **describe** how you found it.
P40 (Challenge)	A learner writes the estimated mean as 13.1666667 goals. **Describe** what is wrong and **state** the answer to 3 significant figures.

Concept 5 · One extreme value

P41 (Fluency)	The nine sponsored-walk amounts are printed above. **State** their mean.

P42 (Fluency)	The nine sponsored-walk amounts are printed above. **State** their median.
P43 (Fluency)	One more amount of 130 pula joins the nine. **Calculate** the new mean.
P44 (Fluency)	One more amount of 130 pula joins the nine. **State** the new median.
P45 (Applied)	**Calculate** how far the mean moved when the 130 pula amount joined.
P46 (Applied)	**Calculate** how far the median moved when the 130 pula amount joined.
P47 (Applied)	**Describe** which average the extra amount moved more, and why.
P48 (Applied)	**Calculate** the range of the nine amounts and of the ten amounts.
P49 (Challenge)	The 130 pula amount is found to belong to another school and is taken out again. **Calculate** the mean of the remaining nine amounts and **describe** what that shows.
P50 (Challenge)	**Determine** what a tenth amount would have to be for the mean of the ten to stay at 40 pula.

Concept 6 · Which average, and why

P51 (Fluency)	**State** which average can be used for a variable whose answers are words, such as favourite colour.
P52 (Fluency)	**State** which average is least affected by one extreme value.
P53 (Fluency)	**State** which average uses every value in the dataset.
P54 (Fluency)	Every value in a list occurs exactly once. **State** which average does not exist for it.

P55 (Applied)	A shop must decide which shoe sizes to order. **Describe** which average it should use and why.
P56 (Applied)	A small workshop reports what a worker there typically earns. One person earns many times what everyone else does. **Describe** which average to use and why.
P57 (Applied)	A class test has marks spread evenly with no extreme values. **Describe** which average to use and why.
P58 (Applied)	A shop says “the average shirt size we sell is 8.4”. **Describe** why that number is not useful.
P59 (Challenge)	The ten walk amounts have mean 49, median 47.5 and mode 50 pula. **Describe** why “the average was 49 pula” is not enough on its own.
P60 (Challenge)	A tuck shop wants to know which drink to order more of next term. **Describe** the average it needs and **state** what data it must collect.

CRITERION B · INVESTIGATING PATTERNS

The newcomer

A list of n values has a mean of exactly **10**. One new value, x, joins the list. The new mean depends on two things: how big x is, and how many values were already there. This investigation is about the second one.

Part 1 · the structured investigation

1. **Predict** what the new mean will be when $n = 9$ and $x = 22$. Write the prediction down before you work anything out.

2. Take $n = 4$ values with a mean of 10 — for example 7, 9, 11, 13 — and **tabulate** the new mean when $x = 10, 16, 22$.

3. **Identify** the pattern in your table. Say what happens to the new mean each time x goes up by 6.

4. **Determine** the general rule for the new mean when n values of mean 10 are joined by one value x, and write it as a formula.

5. **Test and verify** the rule: repeat the whole table for $n = 9$, using a list of 9 values you choose yourself with mean 10, and compare every answer with your formula.

6. **Justify** why the same value of x shifts the mean less when n is larger.

New value, x	New mean when n = 4	New mean when n = 9
10	10	
16		
22		
x		

The first entry is done for you: joining a value of 10 to four values whose mean is already 10 leaves the mean at 10. The rest of the table is yours, and the last row is the general rule.

Part 2 · your own method choice

Part 2 · your own investigation. Investigate the same question for the **median**. A list of n values has a median of 10; one new value x joins it. Predict, tabulate, find the pattern, state a general rule if there is one,

test it and justify what you find. You choose the lists, the values of x, the table and the way you present the result — and be careful: the answer may depend on things the mean did not care about.

Criterion B · Investigating patterns. Levels 1–2: works out some new means correctly for $n = 4$. **Levels 3–4:** completes the table correctly and describes the pattern in words. **Levels 5–6:** states a correct general rule for n values and x, and verifies it against the $n = 9$ table. **Levels 7–8:** justifies the rule from the structure of the mean — the new total is the old total plus x, and the shift is $(x - 10)$ shared between $n + 1$ values, so it shrinks as n grows — and carries Part 2 through to a tested general statement about the median.

Remember, you should communicate in an organized and coherent manner.

CRITERION D · APPLYING MATHEMATICS IN REAL-LIFE CONTEXTS

The average pay at the brickworks

A small brickworks in Serowe employs eleven people. The owner works there too and pays herself out of the same business. An advertisement for a new job is about to say what the **average pay** at the brickworks is, and the wage sheet below is all the information there is. The sheet carries more columns than the question needs.

Role	Monthly pay (P)	Years worked	Shift	Clock no.
Moulder	2550	3	day	101
Kiln loader	3450	7	day	102
Machine operator	3600	5	day	103
Yard hand	2150	1	day	104
Yard hand	2050	2	night	105
Driver	3850	9	day	106
Supervisor	4150	12	day	107
Store clerk	4050	4	day	108
Moulder	2900	6	night	109
Kiln loader	3150	8	night	110
Cleaner	2300	2	day	111
Owner	24000	18	day	100

This wage sheet is an invented illustrative example written for the task. It is not taken from any business, price list or publication, and no organisation, person or source is named or implied.

1. **Interpreting the situation. Describe** which column of the sheet answers “what is the average pay here?”, and which columns are surplus to that question. Say what each surplus column would be useful for instead.

2. **Selecting the mathematics. State** which averages can be worked out from a pay column, and **describe** why the mode is of little use here.

3. **Applying the mathematics. Calculate** the mean pay and the median pay of all 12 people on the sheet. Then **calculate** the mean pay of the eleven workers with the owner's pay left out.

4. **Accuracy. Check** your mean against the total you added, and **state** both averages to the nearest pula. Say how many values you ordered to find the median, and which two you used.

5. **Sense in context. Describe** which average the advertisement would print, and which one honestly describes what a worker there earns. Then **describe** one thing this sheet cannot tell a job applicant.

Criterion D · Applying mathematics in real-life contexts. Levels 1–2: identifies the pay column and makes a start on one average. **Levels 3–4:** selects the right column, calculates the mean correctly and orders the sheet to find the median. **Levels 5–6:** all three figures correct with the total shown, the surplus columns identified, and a described recommendation naming the median as the honest figure. **Levels 7–8:** in addition, describes clearly that the mean of P4850 is higher than any of the eleven workers actually earns, so a single value can make an average describe nobody — and describes something the sheet cannot show, such as hours worked for that pay, which no pay column can ever reveal.

ANSWERS

Every answer here is recomputed from the datasets printed in the questions before this file is written.

P1	10	**P31**	$10 \leq x < 15$
P2	7	**P32**	7.5
P3	9	**P33**	8
P4	16 marks	**P34**	30
P11	12	**P41**	40 pula
P12	8	**P42**	45 pula
P13	8th	**P43**	49 pula
P14	3 and 7	**P44**	47.5 pula
P21	10	**P51**	the mode
P22	20	**P52**	the median
P23	2 correct answers	**P53**	the mean
P24	30	**P54**	the mode

Applied and Challenge · fully worked

P5 12 marks

Add all eleven marks first: the total is 132. There are 11 marks, so the mean is 132 ÷ 11 = 12 marks. The total is worth writing down on its own line, because a mean that is not checked against a total is a mean nobody can mark.

P6 11 marks

Order the eleven marks: 8, 9, 9, 10, 10, 11, 13, 14, 16, 16, 16. With n = 11 the median position is (11 + 1) ÷ 2 = 6, so count in to the 6th value, which is 11 marks. Counting in from either end lands on the same value, which is a quick check.

P7 16.5

The total is 66 and there are 4 values, so the mean is 66 ÷ 4 = 16.5. No one scored 16.5. The mean is what each would have if the total were shared out equally, and an equal share very often lands between two of the actual values.

P8 42

The mean is the total divided by how many values there are, so the total is the mean multiplied by how many: 6 × 7 = 42. Running the mean backwards like this is how nearly every “find the missing value” question is unlocked.

P9 9

Five values with mean 9 must have a total of 5 × 9 = 45. The four known values add to 6 + 8 + 10 + 12 = 36. The fifth value is therefore 45 − 36 = 9. Check by adding all five: 6 + 8 + 9 + 10 + 12 = 45, and 45 ÷ 5 = 9.

P10 the mean uses every value; the median uses only the position of the middle value

The mean is built from the total, so changing any single value changes it. The median is built from the order: it only asks which value is in the middle, so changing the largest value to something much larger leaves the median exactly where it was. That single difference is why the two averages disagree whenever a dataset has an extreme value.

P15 45 pula

In order the nine amounts are 10, 20, 35, 40, 45, 50, 50, 50, 60. With n = 9 the median position is (9 + 1) ÷ 2 = 5, so the median is the 5th amount, 45 pula. Four amounts lie below it and four lie above it, which is the check.

P16 halfway between the 6th and the 7th values

(12 + 1) ÷ 2 = 6.5, so the median position falls between the 6th and the 7th values of the ordered list. The median is the mean of those two values, and with an even number of values it need not be one of the values at all.

P17 14

In order: 8, 9, 11, 14, 17, 22, 30. There are 7 values, so the median position is (7 + 1) ÷ 2 = 4, and the 4th value is 14. Note that the middle of the list *as written* was 11, which is not the median at all.

P18 47.5 pula

In order the ten amounts are 10, 20, 35, 40, 45, 50, 50, 50, 60, 130. With n = 10 the median position is (10 + 1) ÷ 2 = 5.5, so the median is halfway between the 5th and 6th values: (45 + 50) ÷ 2 = 47.5 pula.

P19 14

They took the middle of the list as it was written down. The median is defined on the *ordered* list: 8, 9, 11, 14, 17, 22, 30, whose 4th value is 14. The raw list almost never arrives in order, so ordering it is the first move, not an optional tidy-up.

P20 both are 12

With 8 values the median is the mean of the 4th and 5th values. If those two values are equal, their mean is that value itself, so both must be 12. If they were different — say 11 and 13 — their mean would still be 12, so “equal” is the extra piece of information that pins them both down.

P25 69

Multiply each score by its frequency and add the column: 0 + 6 + 20 + 18 + 20 + 5 = 69. That total is the sum of all 30 learners' scores, without any of them being written out again.

P26 2.3 correct answers

mean = Σfx ÷ Σf = 69 ÷ 30 = 2.3 correct answers. The divisor is the number of learners, 30, and never the number of rows in the table.

P27 2

With 30 values the median position is (30 + 1) ÷ 2 = 15.5, so it lies between the 15th and 16th scores. Counting down the frequency column: 2 learners scored 0, then 6 more reach 8, then 10 more reach 18. Both the 15th and the 16th learner are in that row, so the median is 2.

P28 2 is the value; 10 is only how many learners had it

10 is the largest number in the frequency column, so it is the *frequency* of the mode, not the mode itself. The mode is the score that occurred most often, which is 2 correct answers. An answer to “what is the mode?” is always in the units of the variable.

P29 2.3 correct answers

They divided by 6, the number of *rows*, instead of by Σf = 30, the number of *learners*. The correct mean is 69 ÷ 30 = 2.3. The check is that an average must lie between the smallest and the largest value: 11.5 is not even a possible score out of 5.

P30 12

“3 or more” means the rows x = 3, x = 4 and x = 5, so add those three frequencies: 6 + 5 + 1 = 12 learners. The row x = 3 is included, because “or more” keeps the value itself.

P35 137.5

The midpoint of $10 \leq x < 15$ is $(10 + 15) \div 2 = 12.5$, and its frequency is 11, so $f \times m = 11 \times 12.5 = 137.5$. Every class contributes its midpoint that many times.

P36 395

Multiply each midpoint by its frequency and add: $5 + 45 + 137.5 + 140 + 67.5 = 395$. That total is the estimated total number of goals scored by all 30 players.

P37 13.2 goals

estimated mean $= \Sigma fm \div \Sigma f = 395 \div 30 = 13.2$ goals to 3 significant figures. The word *estimated* belongs in the answer: the table no longer holds the players' real goal counts.

P38 the raw values are gone, so every value in a class is assumed to sit at the midpoint, which makes the answer an estimate

Grouping threw the individual goal counts away. To get a total at all you have to assume every player in a class scored the class midpoint, and almost none of them did. The arithmetic is exact; the assumption is not, so the answer is an estimate and is written as one.

P39 $10 \leq x < 15$

The median position is $(30 + 1) \div 2 = 15.5$, so it lies between the 15th and 16th players. Counting down the frequency column: 2, then 6 reaches 8, then 11 reaches 19 — so both the 15th and the 16th player are in the class $10 \leq x < 15$. The table cannot say where in that class the median sits, only which class holds it.

P40 13.2 goals

Nine figures of accuracy from a table that no longer holds the real goal counts is a false claim of precision. The division $395 \div 30$ does not stop, so the answer is rounded: 13.2 goals to 3 significant figures. Rounding an estimate is not sloppiness; quoting it unrounded is dishonesty about what the table knows.

P45 9 pula

The mean was 40 pula and is now 49 pula, so it moved $49 - 40 = 9$ pula. The new total is 490 and there are now 10 amounts, which is where the new mean comes from.

P46 2.5 pula

The median was 45 pula and is now 47.5 pula, so it moved 2.5 pula. The new value only pushed the middle position along by half a place: it is now halfway between the 5th and 6th amounts instead of on the 5th.

P47 the mean, because it is built from the total

The mean moved 9 pula and the median only 2.5 pula. Every value feeds into the mean through the total, so a value far from the rest pulls it hard. The median only cares about position, so one extra value at the far end slides it along the order by half a place and no further.

P48 50 pula and 120 pula

Range = largest − smallest. Before: 60 − 10 = 50 pula. After: 130 − 10 = 120 pula. One value more than doubled the range while the median barely moved — a first look at why S07 measures spread separately from centre.

P49 40 pula

Removing it takes the total back to 360 over 9 amounts, so the mean returns to exactly 40 pula. That is the point: the whole 9 pula shift was the work of one value out of ten. An average that one entry can move that far should always be quoted with the size of the dataset beside it.

P50 40 pula

Ten amounts with a mean of 40 need a total of 10 × 40 = 400. The nine amounts already total 360, so the tenth must be 400 − 360 = 40 pula — the mean itself. Adding a value equal to the mean never moves the mean, which is the cleanest way to see what the mean actually balances.

P55 the mode

The shop wants the size it will sell most of, which is the mode. The mean shoe size could easily come out as 7.4, and no shop can order a size 7.4 shoe. When the question is “which one is most popular?”, the mode is the only average that answers it.

P56 the median

The one very large pay drags the mean up to a figure nobody actually earns. The median ignores how far out that value is and reports the middle worker, so it still describes a typical pay. The mean is not wrong arithmetic — it is simply the wrong question.

P57 the mean

With no extreme values there is nothing to distort the total, so the mean is safe and it has the advantage of using every mark. The median would throw away most of the information the class produced.

P58 shirt sizes are counted in fixed steps, so 8.4 is not a size; the mode is what the shop needs

The shop has quoted a mean of a discrete variable. There is no size 8.4 to order, and the number does not tell the shop which sizes to stock. The mode — the size that sells most — is the average this question wants.

P59 it does not say which average, and the three disagree

Three averages, three different numbers: 49, 47.5 and 50 pula. Quoting “the average” without naming it lets the writer pick whichever number suits them — here the mean is the largest, and it is the one a fundraising poster would print. Naming the average, and saying how many values it came from, is the honest form.

P60 the mode; a record of which drink each sale was

“Which drink” is a category, so no mean or median exists — only the mode. What the shop must collect is one row per sale naming the drink, which is the tally-and-frequency-table job of S01. The most common drink in that table is the answer, and no other average is available for words.