13 CHAPTER 13 Statistics

WHAT THIS CHAPTER BUILDS
You will learn to find the mean, the median and the mode of grouped data, and which of the three to trust when one value sits far from the rest.

STATISTICS AND PROBABILITY UNIT

IN THIS CHAPTER

Meera and Arjun sort a heap of potatoes into four crates, one group to a crate.

Getting started

Summarising grouped data

A STORY

Four crates, not two hundred potatoes

Meera lifts a potato from the heap and weighs it in her hand. Arjun drops another into a crate.

“We could weigh every potato,” Arjun says, “but there must be two hundred of them.”

“So we sort them instead,” Meera says. “Under 100 grams in one crate, 100 to 200 in the next, then 200 to 300, then 300 to 400.”

“Then we only count how many land in each crate,” Arjun says. “Four numbers, not two hundred.”

“But then we lose each potato’s exact weight,” Meera says. “How do we find the average?”

“Treat every potato in a crate as if it weighed the middle of its range,” Arjun says. “50 grams, 150, 250, 350. The answer comes out close.”

“And the fullest crate tells us where the most common size lies,” Meera says.

Arjun looks along the four crates. “So four counts can still give us the average, the middle and the most common.”

You will learn to find the mean, the median and the mode when the data comes in groups like these.

Data often arrives already grouped into class intervals, not as single values. A wage survey might report how many workers fall in ₹200–250, then ₹250–300, and so on. Weights, ages and distances get grouped the same way.

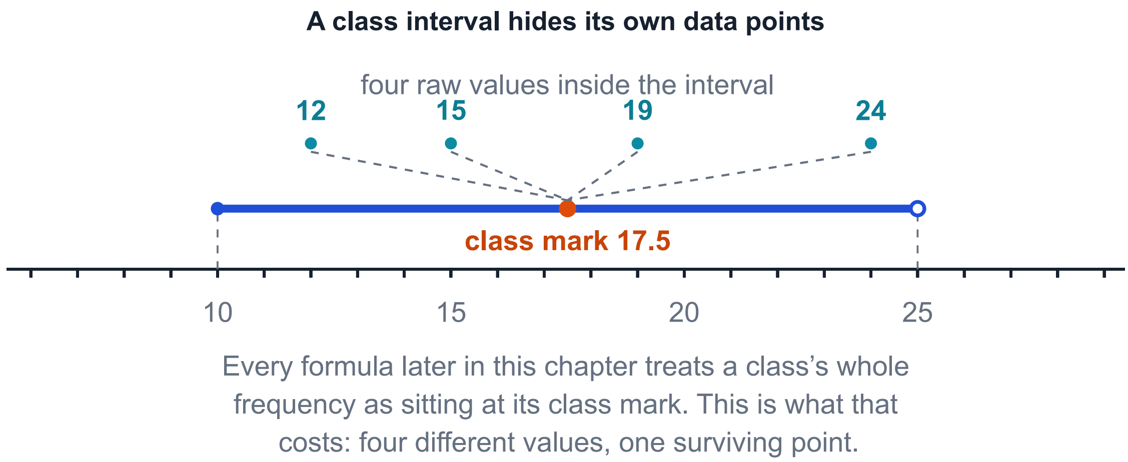
This chapter finds the mean, the median and the mode of grouped data. The mean has three methods: direct, assumed-mean and step-deviation. We pick the method that suits the numbers in front of us.

All three methods give the same mean on the same frequency table. The median and the mode each use one formula instead, keyed to one specific class.

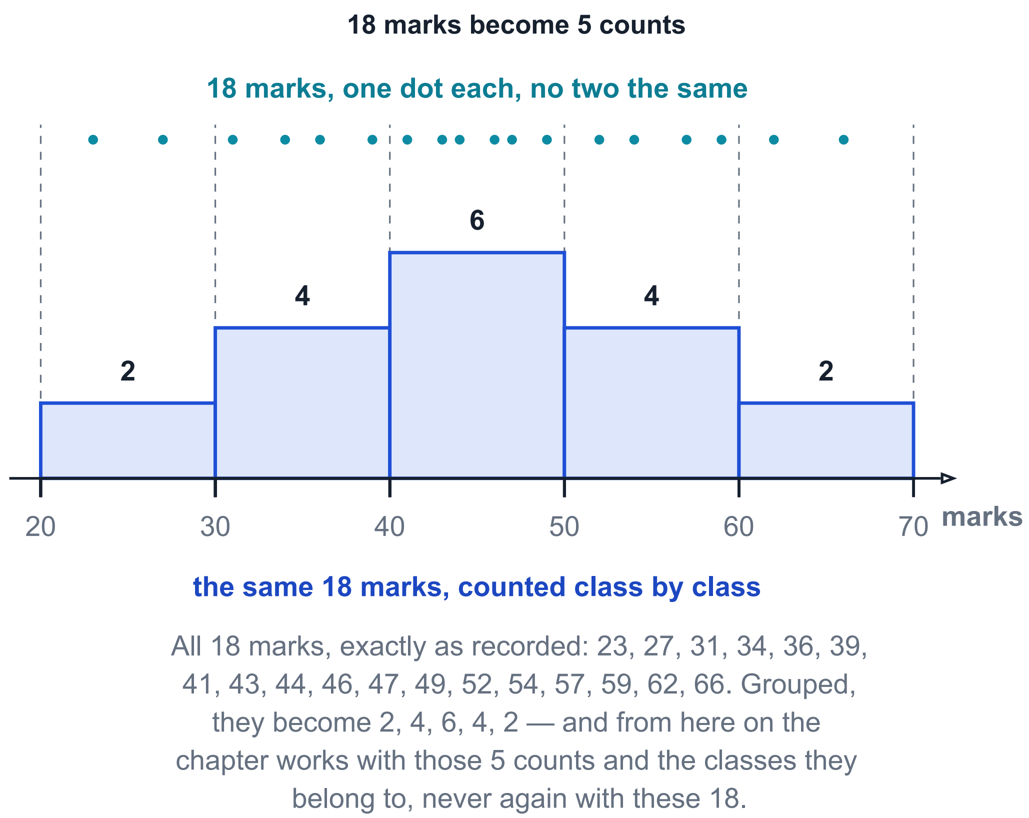
Before opening any formula, read the modal class or the median class from the frequency column first. That one step decides which class’s numbers go into the formula. Get this step wrong, and every later number in the solution is wrong too.

RULE
Identify the modal class or the median class from the frequency column before opening either formula.

KEY
A class interval hides its own data points, so mean, median and mode here come from frequencies, not raw values.



Four raw values, 12, 15, 19 and 24, sit inside one class interval, and the class mark 17.5 stands in for all of them.



Grouping keeps all eighteen marks in the count but throws away where each one sat, leaving only five counts to work with.

A SCHOLAR INDIA REMEMBERS

Dignaga was born in south India, around the 5th or 6th century CE. He wrote a book called the Pramanasamuccaya. In it he asked what really counts as a way of knowing. He kept only two. One is what you see. The other is what you work out from what you see. A table of grouped data needs the same care. Some numbers you counted. Others, like a class mark, you worked out.

Dignaga · the kingfisher

IN THE WILD

A school's result sheet often gives only how many students scored in each ten-mark band (12 students between 40 and 50, say), never anyone's exact mark. The class average has to come from the band counts alone, and this method finds it from that table, without a single individual score.

CHECK YOURSELF

1. This chapter's methods for the mean, median and mode of grouped data all start from

- **A.** the individual raw values, listed one by one, exactly as in ungrouped data
- **B.** a frequency table of class intervals, not the individual raw observations
- **C.** only the single class with the highest frequency
- **D.** a graph of the data, plotted before any frequency table is made

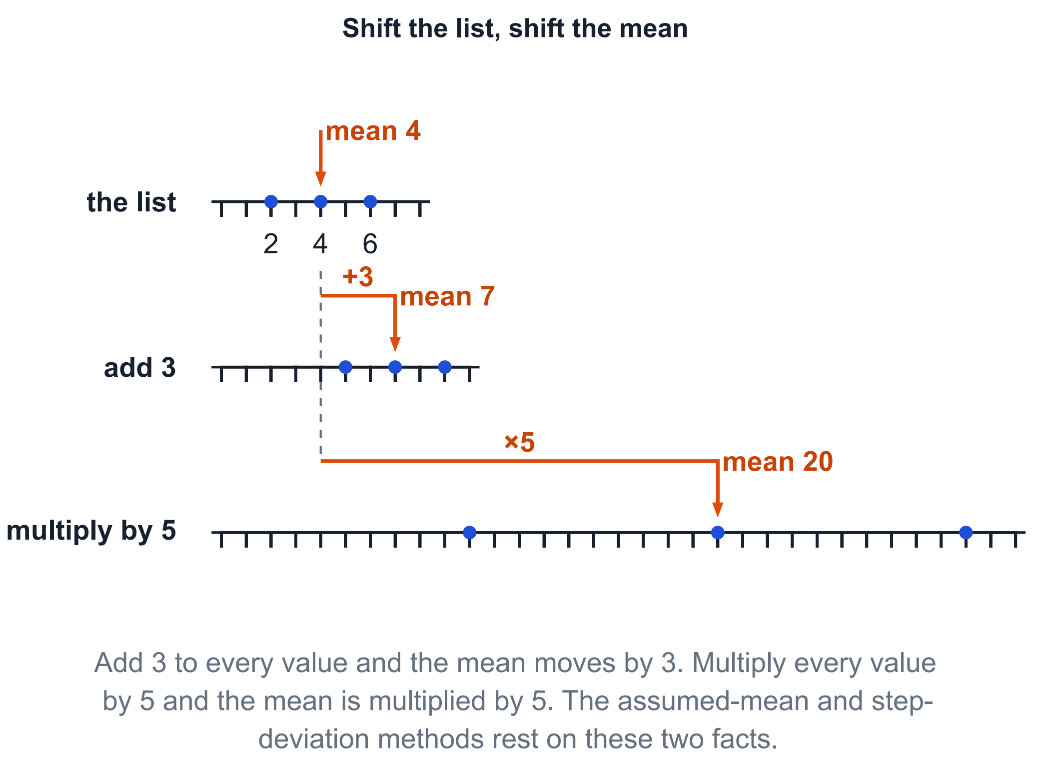
2. Why can the mean, median and mode of grouped data be computed without knowing the individual values inside each class?

- **A.** because every value inside a class is actually identical to that class's mark
- **B.** because the total number of observations, n, is never needed once data is grouped
- **C.** a class's frequency is treated as concentrated at one point
- **D.** because grouped data always has classes of equal size

1 more in this set online.

IN EVERY CHAPTER

Before you start



Adding 3 to every value shifts the mean from 4 to 7, and multiplying by 5 scales it from 4 to 20.

DO THIS FIRST

You already found the mean and median of a plain list. Now do the same job on grouped data. Try each check below.

- **Mean of a list** (Class 8, Ganita Prakash Part 2, page 107).
  Here: the direct method finds it by using each class's mark as x_i.
  Check: The mean of $2, 4, 6$ is $(2+4+6)/3 = 4$.

- **Median of a list** (Class 8, Ganita Prakash Part 2, page 108).
  Here: the median class holds the middle-ranked value, once grouped.
  Check: The median of $1, 3, 5, 7, 9$ is 5.
- **Average of two numbers** (Class 8, Ganita Prakash Part 2, page 109).
  Here: the class mark averages a class's two limits, lower and upper.
  Check: The average of 8 and 11 is $(8 + 11)/2 = 9.5$.
- **Adding a fixed number** (Class 8, Ganita Prakash Part 2, page 106).
  Here: the assumed-mean method shifts every value by a fixed number, a.
  Check: Adding 3 to each of $2, 4, 6$ shifts the mean from 4 to 7.
- **A weighted average** (Class 8, Ganita Prakash Part 2, page 110).
  Here: the direct method weights each class mark by its own frequency.
  Check: Values $4, 6$ with frequencies $3, 2$ average to 4.8.
- **Scaling every value** (Class 8, Ganita Prakash Part 2, page 108).
  Here: the step-deviation method undoes this scaling, by a class size h.
  Check: Multiplying each of $2, 4, 6$ by 5 scales the mean to 20.

If any of these felt new, read the page named before going on.

The class mark of a class interval

The class mark of a class interval is the average of its lower and upper limits. For the interval 10–25, lower limit 10 and upper limit 25 give class mark $(10 + 25)/2 = 17.5$.

Every grouped-data formula treats a class's frequency as sitting at this one point. No formula uses a class's lower and upper limits on their own. Once we have found the class mark, it is the number we track.

Economics reports household income, and Geography reports rainfall or crop yield, in class intervals such as ₹10,000–₹20,000. Both subjects use the class mark when they compute an average.

Your turn: what is the class mark of the interval 60–75? (Answer: 67.5, since $(60 + 75)/2 = 67.5$.)

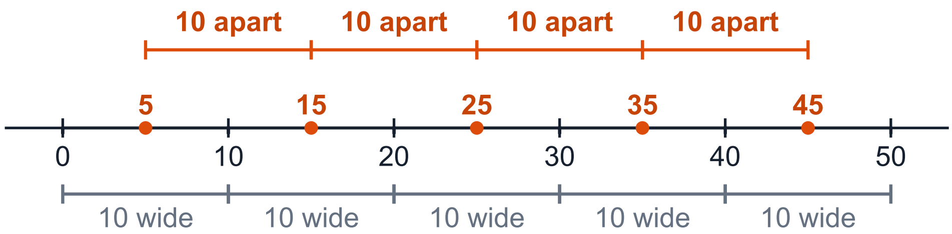
TRY
Class mark of the interval 40–55?

Answer: $(40 + 55)/2 = 47.5$.

BACK
Summarising grouped data · p. 433

KEY
The class mark is the midpoint of a class interval: average its lower and upper limits.

The class marks make a ruler of their own



Each mark is the average of its own class's two limits, so it stands half a class above the lower one: $(0+10)/2 = 5$, then $(10+20)/2 = 15$. That one fact makes the marks a copy of the boundary ruler slid along by 5, and it is why neighbouring marks are always 10 apart and never anything else.

Each class mark sits half a class above its own lower limit, so the marks form a second ruler shifted along from the boundaries.

A SCHOLAR INDIA REMEMBERS

What does the class mark stand for? Say it exactly. It is the middle value of a class. For the class 10 *to* 20*, it is* $(10+20)/2 = 15$*. We then treat every value in that class as if it were* 15*. That is an assumption. Know that you made it.*

Dignaga · the kingfisher

IN THE WILD

Economics reports household income, and Geography reports rainfall or crop yield, in class intervals such as ₹10,000–₹20,000, because there are too many raw values to list one by one. The class mark is the number that stands for each interval when an average is found.

CHECK YOURSELF

3. The class mark of the interval 24-36 is

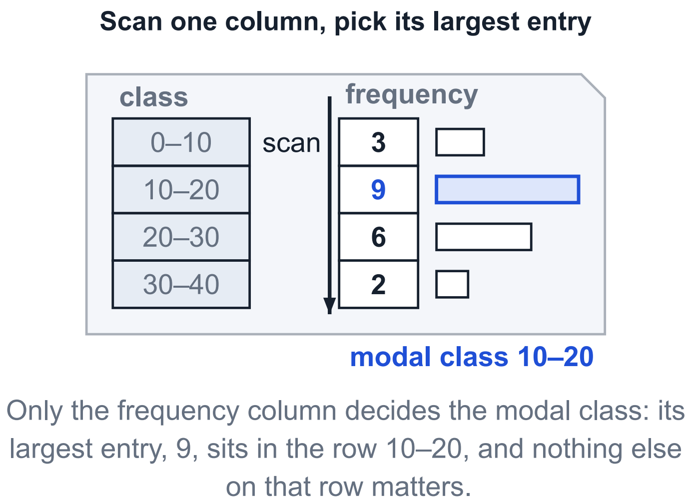
- **A.** $36 - 24 = 12$
- **B.** $(24+36)/2 = 30$
- **C.** 24, the lower limit of the class
- **D.** 36, the upper limit of the class

4. A table lists the class 45-65 with frequency 8. In the direct-method mean formula, which single number represents every one of those 8 observations?

- **A.** the frequency, 8
- **B.** the class width, 20
- **C.** the lower limit, 45
- **D.** the class mark, 55

1 more in this set online.

The modal class



The modal class is read from one column alone, so a class with big values but a small count can never be it.

The class interval with the highest frequency in a grouped table is the modal class. Nothing more decides it: only the frequency column matters. Scan the frequency column: the modal class is the number with the biggest count.

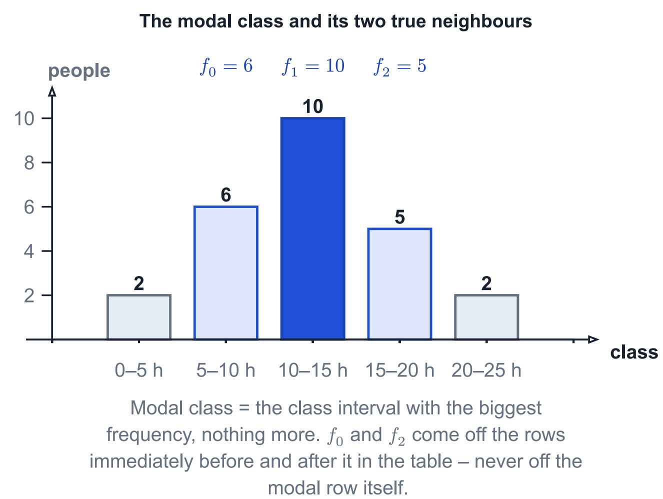
A distribution can have two classes tied for the highest frequency. That tied case is called bimodal, and it does not appear on these pages.

No other column in the table plays any part in identifying the modal class. Once that one number is found, nothing else on the row matters.

Your turn: classes 0–10, 10–20, 20–30 and 30–40 have frequencies $3, 9, 6, 2$. Which is the modal class? (Answer: 10–20, since 9 is the highest frequency.)

CAUTION
A tie for highest frequency is a genuinely harder case (multimodal), avoided here.

KEY
Modal class = the class interval with the biggest frequency, nothing more.



The tallest bar marks the modal class, with its two neighbouring bars shaded to show the frequencies the mode formula needs.

CHECK YOURSELF

5. In a grouped frequency table, the modal class is

- **A.** the class interval listed first in the table
- **B.** the class interval containing the largest data values
- **C.** the middle class interval in the table
- **D.** the class interval with the highest frequency

6. For the ages of 30 people grouped as 5-15, 15-25, 25-35, 35-45, 45-55 with frequencies 3, 5, 9, 7, 6, the modal class is

- **A.** 25-35
- **B.** 45-55
- **C.** 5-15
- **D.** 35-45

1 more in this set online.

The median class

The median class is the class interval containing the middle-ranked observation. To find it, we add frequencies class by class, in the given order, keeping a running total as we go.

The median class is the first class where that running total reaches or passes half the total observations, $n/2$. No later class can be the median class, even if its own frequency is bigger. The class holding the running total the moment it first reaches $n/2$ is the median class.

For example, with $n = 60$, the median class is the first class whose running total reaches at least 30. The same running total method works on any table, regardless of what n turns out to be.

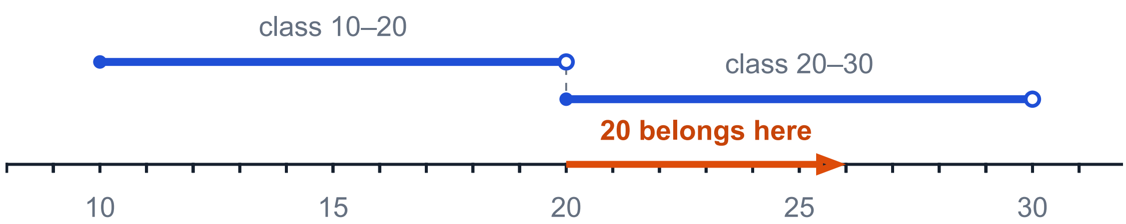
Your turn: with $n = 80$, what running total first makes a class the median class? (Answer: the first class whose running total reaches 40, since $n/2 = 80/2 = 40$.)

TRY
If $n = 60$, what running total marks the median class?

Answer: The first class where the running total is ≥ 30.

KEY
Add frequencies class by class until the running total first reaches $n/2$: that class is the median class.

Which class does the boundary 20 belong to?



A continuous class runs from its lower limit up to, but not including, its upper limit. The value 20 is the upper limit of 10–20 and the lower limit of 20–30 at once – only one of those two facts includes it.

A filled dot marks where a class starts, an open circle where it ends, so the boundary 20 belongs to only one class.

A SCHOLAR INDIA REMEMBERS

Modal class and median class sound alike. Are they the same kind of thing? No. The modal class has the highest frequency. The median class holds the middle value, and we find it from the cumulative frequency. In one table they can be two different classes.

Dignaga · the kingfisher

CHECK YOURSELF

7. The median class of grouped data is located by

A. the first class where the running total reaches $n/2$

B. finding the class with the highest frequency

C. finding the middle class by position in the table, regardless of frequency

D. finding the class whose class mark is closest to the overall mean

8. 68 boxes are weighed and grouped 40-45, 45-50, 50-55, 55-60, 60-65 kg with frequencies 8, 14, 20, 15, 11. With $n/2 = 34$, the median class is

A. 45-50

B. 50-55

C. 55-60

D. 60-65

1 more in this set online.

The ideas

Mean of ungrouped data

The mean of n observations is their total divided by their count. For observations $x_1, x_2, \ldots, x_n$, the mean is $\overline{x} = (\sum x_i)/n$: add everything up, then divide by how many numbers there are.

For example, the mean of $4, 6, 8, 10$ is $(4 + 6 + 8 + 10)/4 = 7$.

Sometimes a value repeats. If a value x_i occurs with frequency f_i, the same idea still applies: it reads $\overline{x} = (\sum f_i x_i)/(\sum f_i)$, each value counted as many times as it occurs.

Grouped data only changes what counts as x_i, never the total-divided-by-count idea behind the mean.

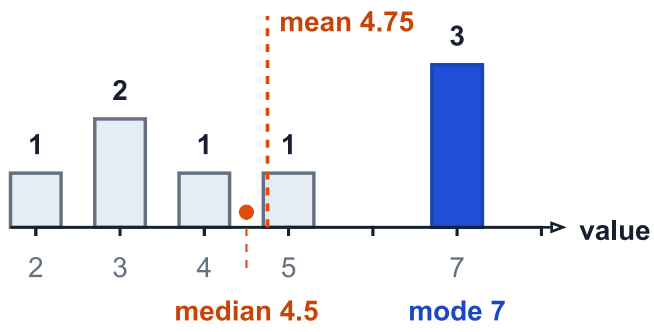
Your turn: what is the mean of $5, 7, 9, 11$? (Answer: 8, since $(5 + 7 + 9 + 11)/4 = 32/4 = 8$.)

TRY
Mean of $4, 6, 8, 10$?

Answer: $(4 + 6 + 8 + 10)/4 = 7$.

KEY
Mean = total of all values divided by how many values there are.

Mean, median, mode: three answers, one dataset



Eight values, none of them grouped: $2, 3, 3, 4, 5, 7, 7, 7$.
The mean and the median both land between bars; the mode is a bar the data actually has.

Eight ungrouped values give a mean of 4.75, a median of 4.5 and a mode of 7, each at a different point.

CHECK YOURSELF

9. The mean of the 5 numbers 12, 15, 18, 20, 25 is

- **A.** $90/5 = 18$
- **B.** 90, the sum, left undivided by the count
- **C.** 18.5, the average of the smallest and largest values only
- **D.** 5, the count of numbers, mistaken for the mean

10. Why does $\overline{x} = (\sum f_i x_i)/(\sum f_i)$ give the same result as adding every value once and dividing by the count, when a value x_i repeats f_i times?

- **A.** summing $f_i x_i$ adds x_i to itself f_i times, matching the total
- **B.** because f_i and x_i always turn out equal for repeated values
- **C.** because dividing by $\sum f_i$ instead of the raw count gives a more accurate answer
- **D.** because $f_i x_i$ rounds the value x_i to the nearest frequency

1 more in this set online.

Median of ungrouped data

To find the median of ungrouped data, we first arrange the observations in ascending order, smallest to largest. The median is the value sitting at the middle position.

If the count n is odd, one value sits exactly in the middle: the median is the value at position $(n + 1)/2$. An odd count always leaves exactly one middle value.

If n is even, two values share the middle. **The median is then the average of the values at positions $n/2$ and $n/2 + 1$.** An even count always averages the two values sitting either side of the middle.

For example, the ordered data $3, 7, 9, 12$ has $n = 4$, even. The median averages positions 2 and 3: $(7 + 9)/2 = 8$.

Your turn: what is the median of $2, 5, 6, 8, 9$? (Answer: 6, since $n = 5$ is odd and 6 sits at position $(5 + 1)/2 = 3$.)

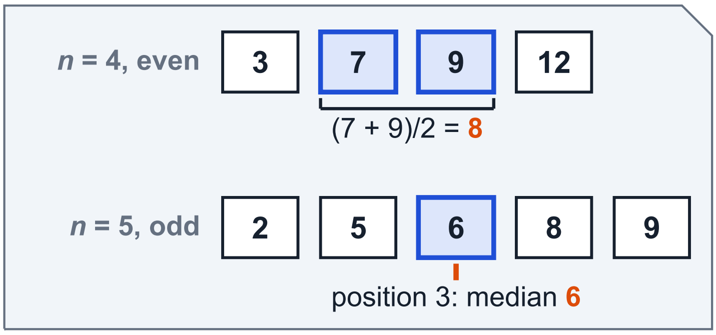
BACK
Summarising grouped data · p. 433

TRY
Median of the ordered data $3, 7, 9, 12$?

Answer: Even count, so average positions 2 and 3; $(7+9)/2 = 8$.

KEY
Odd count: one middle value. Even count: average the two middle values.

Odd count, one middle; even count, two



With four values two share the middle, so the median averages 7 and 9 to 8; with five values one sits in the middle, position 3, and the median is 6.

The median averages the middle two values, 7 and 9, to 8 for four values, or is the middle value, 6, for five.

CHECK YOURSELF

11. Arranged in order, 8, 11, 15, 19, 23 has $n = 5$ (odd). The median is

- **A.** 8, the smallest value
- **B.** 23, the largest value
- **C.** $(8+23)/2 = 15.5$, the average of the smallest and largest
- **D.** 15, the value at position $(5+1)/2 = 3$

12. For 9, 13, 17, 25, 31, 35 (already in order, $n = 6$, even), the median is

- **A.** 17, the value at position $n/2 = 3$ alone
- **B.** 25, the value at position $n/2 + 1 = 4$ alone
- **C.** $(17+25)/2 = 21$
- **D.** $(9+35)/2 = 22$, the average of the smallest and largest

1 more in this set online.

Mode of ungrouped data

The mode of ungrouped data is the value that occurs most often: the observation with the highest frequency. We count how many times each value shows up, and the mode is the value with the highest count.

For example, in the data $2, 3, 3, 5, 3, 7$, the value 3 occurs three times: more than any other value. So the mode is 3.

Unlike the mean or the median, the mode need not be unique. Data with two equally frequent values is called bimodal. Bimodal data does not appear on these pages, but the name is worth knowing for when two values tie.

Your turn: what is the mode of $4, 6, 6, 9, 6, 2$? (Answer: 6, since it occurs three times, more than any other value.)

TRY
Mode of $2, 3, 3, 5, 3, 7$?

Answer: 3; it occurs three times, more than any other value.

KEY
Mode = the most frequent value, nothing more.

CHECK YOURSELF

13. For the values 7, 3, 7, 9, 7, 3, 5, the mode is

- **A.** 3, since it is the smallest value that repeats
- **B.** 9, the largest value in the list
- **C.** 7, the most frequent value
- **D.** the mean of all the values

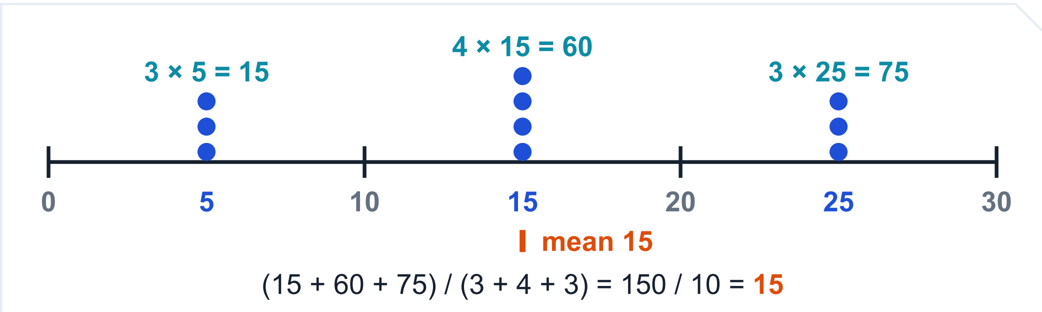
14. Data can be bimodal, but this chapter's mode formula does not extend to that case. Why not?

- **A.** because bimodal data cannot occur in real measurements
- **B.** because the formula always averages the two most frequent values instead
- **C.** because bimodal data has no mean or median either
- **D.** the formula needs one class with the strictly highest frequency

1 more in this set online.

Mean of grouped data: the direct method

Every observation stands at its class mark



The three observations of 0–10 all stand at 5, the four of 10–20 at 15 and the three of 20–30 at 25; the products 15, 60 and 75 add to 150, and 150 over 10 is 15.

Three classes stand at their class marks, 5, 15 and 25, weighted 3, 4 and 3, giving a mean of 15.

The direct method finds the mean of grouped data straight from the class marks and frequencies. The formula is $\overline{x} = (\sum f_i x_i)/(\sum f_i)$, summed over every class in the table. Every number this formula needs is already sitting in the frequency table.

Each class's own class mark serves as its x_i, paired with that class's own frequency f_i. We build a small table, class mark next to frequency, before multiplying a single pair.

Use the class mark for x_i, never a class's boundary numbers on their own. If we ever reach for the boundary numbers instead, the fix is to stop and go back to the class mark.

Your turn: classes 0–10, 10–20 and 20–30 have frequencies $3, 4, 3$ and class marks $5, 15, 25$. What is the mean by the direct method? (Answer: 15, since the products add to 150 and the frequencies add to 10, so $150/10 = 15$.)

RULE
Use the CLASS MARK for x_i, never the class's boundary numbers.

BACK
Mean of ungrouped data · p. 440

KEY
Direct method: plug each class mark and frequency straight into $\overline{x} = (\sum f_i x_i)/(\sum f_i)$.

CHECK YOURSELF

15. By the direct method, the mean of grouped data is computed as

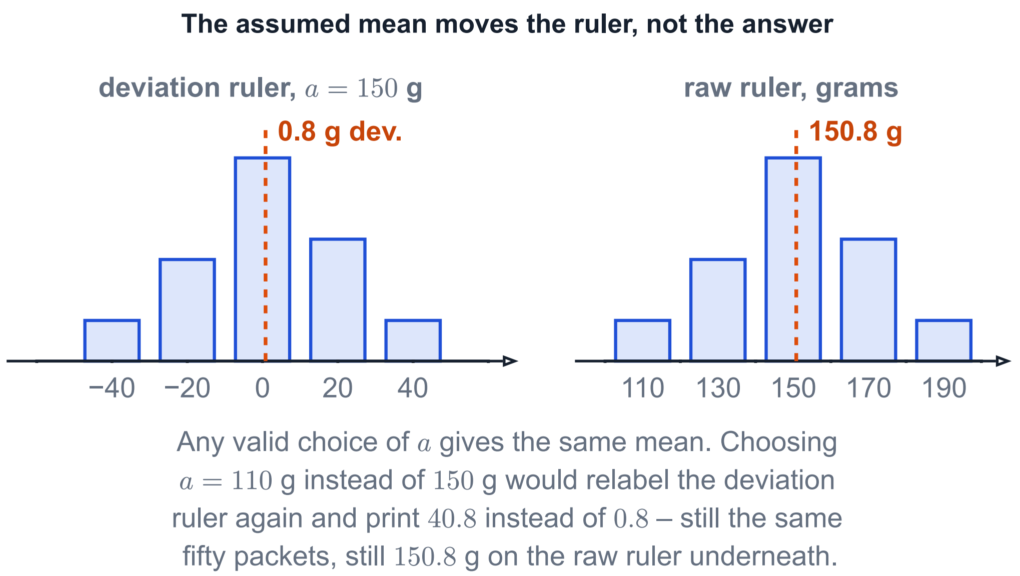
- **A.** $\overline{x} = (\sum f_i)/(\sum x_i)$, with frequencies and marks swapped
- **B.** $\overline{x} = (\sum x_i)/n$, ignoring the frequencies entirely
- **C.** $\overline{x} = (\sum f_i x_i)/(\sum f_i)$, using each class's mark x_i
- **D.** $\overline{x} = (\sum f_i l_i)/(\sum f_i)$, using each class's lower limit instead of its mark

16. The runs of 25 batsmen are grouped as 0-20, 20-40, 40-60, 60-80, 80-100 with frequencies 3, 7, 9, 4, 2. By the direct method, the mean is

- **A.** 36, from using each class's lower limit instead of its mark
- **B.** 1150, the value of $\sum f_i x_i$, left undivided
- **C.** 50, the class mark of the middle class alone
- **D.** 46

2 more in this set online.

Mean of grouped data: the assumed-mean method



The same fifty packets are measured on two rulers, one in grams and one in deviations, with the mean at the same spot on both.

When the class marks are large, working with them directly gets clumsy: it means multiplying big numbers for no real gain. The assumed-mean method picks one class mark as an assumed mean, a, and works with deviations instead.

Set $d_i = x_i - a$ for every class. Then $\overline{x} = a + (\sum f_i d_i)/(\sum f_i)$: the same mean as the direct method, from smaller numbers. We pick a near the middle of the class marks, and the deviations stay small on both sides.

Any valid choice of a gives exactly the same mean. On the tea-packet table, $a = 150$ and $a = 110$ both give 150.8 grams.

Your turn: classes with marks $10, 20, 30$ and frequencies $2, 5, 3$ take $a = 20$. What is the mean? (Answer: 21, since the deviations are $-10, 0, 10$, their weighted total is 10, and $20 + 10/10 = 21$.)

RULE
Any valid choice of a gives the same mean: the method is a shortcut, not a different answer.

KEY
Pick any class mark as a; work with the smaller deviations $d_i = x_i - a$ instead of the class marks themselves.

CHECK YOURSELF

17. The assumed-mean method rewrites the mean as

- **A.** $\overline{x} = a - (\sum f_i d_i)/(\sum f_i)$, subtracting the correction instead of adding it
- **B.** $\overline{x} = (\sum f_i d_i)/(\sum f_i)$, dropping a entirely
- **C.** $\overline{x} = a + (\sum d_i)/(\sum f_i)$, leaving the frequency out of the numerator
- **D.** $\overline{x} = a + (\sum f_i d_i)/(\sum f_i)$, where $d_i = x_i - a$

18. Daily wages of 40 workers are grouped 100-120, 120-140, 140-160, 160-180, 180-200 with frequencies 5, 9, 15, 7, 4. Taking $a = 150$ gives deviations $d_i = -40, -20, 0, 20, 40$ and $\sum f_i d_i = -80$. The mean is

- **A.** 150, just the assumed mean itself
- **B.** 148
- **C.** -2, the correction term alone, without adding back a
- **D.** 152, from adding the correction with the wrong sign

2 more in this set online.

Mean of grouped data: the step-deviation method

Sometimes every deviation d_i shares a common factor: the class size, h. The step-deviation method divides that factor out before multiplying, so the numbers to multiply are far smaller.

Set $u_i = (x_i - a)/h$ for every class. Then $\overline{x} = a + h \cdot (\sum f_i u_i)/(\sum f_i)$: the smallest numbers of the three methods to multiply by hand. We check the deviations for a shared factor before dividing: it is not always there.

Otherwise, the assumed-mean method is just as easy.

Your turn: classes with marks $20, 40, 60$ and frequencies $3, 7, 2$ take $a = 40$ and $h = 20$. What is the mean? (Answer: ≈ 38.33, since $u_i = -1, 0, 1$, their weighted total is -1, and $40 + 20 \cdot (-1/12) \approx 38.33$.)

RULE
Step-deviation only helps when every d_i shares a common factor; otherwise assumed-mean is just as easy.

KEY
Divide each deviation by the class size h first: $u_i = (x_i - a)/h$. Then use the assumed-mean shortcut.

CHECK YOURSELF

19. The step-deviation method sets $u_i = (x_i - a)/h$ and computes the mean as

A. $\overline{x} = a + (\sum f_i u_i)/(\sum f_i)$, without multiplying by h at the end

B. $\overline{x} = a \cdot h + (\sum f_i u_i)/(\sum f_i)$, multiplying a by h instead of the correction term

C. $\overline{x} = h + a \cdot (\sum f_i u_i)/(\sum f_i)$, with the roles of a and h swapped

D. $\overline{x} = a + h \cdot (\sum f_i u_i)/(\sum f_i)$

20. Monthly bills (in hundreds of rupees) for 68 households are grouped 0-20, 20-40, 40-60, 60-80, 80-100. Taking $a = 50$, $h = 20$ gives $u_i = -2, -1, 0, 1, 2$ and $\sum f_i u_i = 34$. The mean is

A. 60

B. 50.5, from leaving out the multiplication by h

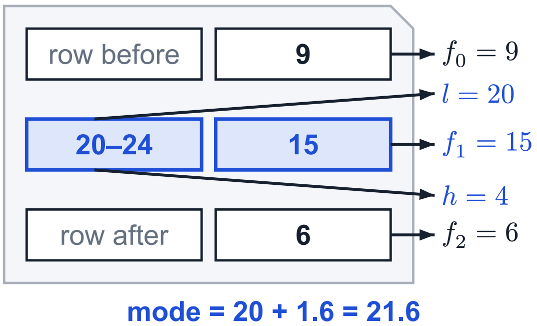
C. 50, the assumed mean before any correction

D. 730, adding a to h times the raw total, skipping the division by $\sum f_i$

2 more in this set online.

Mode of grouped data: the formula

Five numbers off three rows



l and h come off the modal row's class, f_1 off its frequency, and f_0 and f_2 off the rows just above and below: 20 + ((15 - 9)/(30 - 9 - 6)) × 4 = 21.6.

Nothing in the mode formula is worked out from scratch: all five inputs are read straight off three neighbouring rows of the table.

The mode of grouped data comes from one formula, applied to the modal class. It is $l + ((f_1 - f_0)/(2f_1 - f_0 - f_2)) \cdot h$. Five numbers go into this one formula, and every one of them comes from the frequency table.

Here l is the modal class's lower limit and h is the class size. f_1 is the modal class's own frequency. f_0 and f_2 are its two neighbours' frequencies: the classes just before and after it. We find these five numbers on the table before substituting a single one into the formula.

Get f_0 and f_2 right and the final substitution is just arithmetic.

For example, take $f_1 = 8$, $f_0 = 7$ and $f_2 = 2$, with class size 2 and lower limit 3. Then mode $= 3 + ((8 - 7)/(16 - 7 - 2)) \cdot 2 \approx 3.286$.

Your turn: a modal class has $f_1 = 15$, $f_0 = 9$, $f_2 = 6$, class size 4 and lower limit 20. What is the mode? (Answer: 21.6, since $20 + ((15 - 9)/(30 - 9 - 6)) \cdot 4 = 20 + 1.6 = 21.6$.)

TRY
Modal-class frequency 8, preceding class 7, succeeding class 2, class size 2, lower limit 3. Mode?

Answer: $3 + ((8 - 7)/(16 - 7 - 2)) \cdot 2 = 3 + (1/7) \cdot 2 \approx 3.286$.

BACK
Mode of ungrouped data · p. 442

RULE
Read f_0 and f_2 from the table rows before and after the modal class: never recompute them.

CHECK YOURSELF

21. The mode of grouped data is computed as

- **A.** $l + ((f_1 - f_2)/(2f_1 - f_0 - f_2)) \cdot h$, with f_0 and f_2 swapped in the numerator
- **B.** $((f_1 - f_0)/(2f_1 - f_0 - f_2)) \cdot h$, without adding l
- **C.** $l + ((f_1 - f_0)/(f_1 - f_0 - f_2)) \cdot h$, missing the factor of 2 on f_1 in the denominator
- **D.** $l + ((f_1 - f_0)/(2f_1 - f_0 - f_2)) \cdot h$, for the modal class

22. For ages grouped 5-15, 15-25, 25-35, 35-45, 45-55 with frequencies 3, 5, 9, 7, 6, the modal class is 25-35 with $l = 25$, $h = 10$, $f_1 = 9$, $f_0 = 5$, $f_2 = 7$. The mode is

- **A.** 28.33, from swapping f_0 and f_2 in the formula
- **B.** 6.67, from leaving out l
- **C.** 30, the class mark of the modal class, mistaken for the mode
- **D.** 31.67

2 more in this set online.

Median of grouped data: the formula

The median of grouped data comes from one formula, applied to the median class. It is $l + ((n/2 - C)/f) \cdot h$. Five numbers go into this formula, and four of them are already on the frequency table.

Here l is the median class's lower limit and n is the total number of observations. C is the running total of frequencies before that class. f is the median class's own frequency, and h is the class size. We build the running total, class by class, and stop the moment it first reaches $n/2$.

Get C wrong and every later number in the solution shifts too.

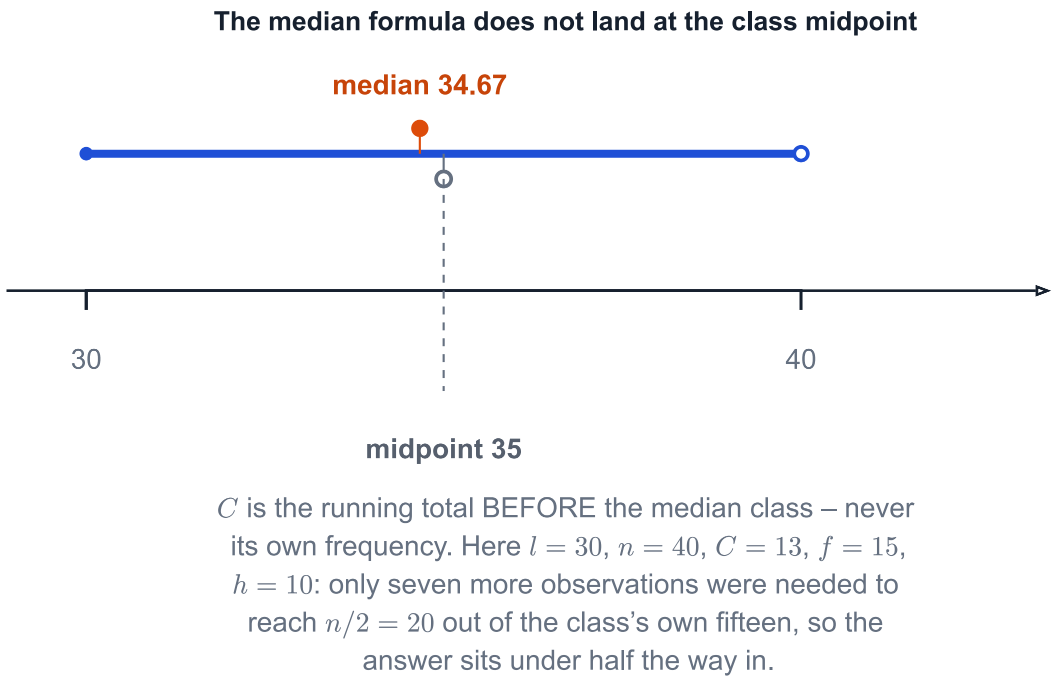
This formula assumes every class has the same size, h, throughout the table. We check our own table for this before trusting the formula on it.

Your turn: a median class has $l = 15$, $n = 50$, $C = 18$, $f = 10$ and $h = 5$. What is the median? (Answer: 18.5, since $n/2 = 25$ and $15 + ((25 - 18)/10) \cdot 5 = 15 + 3.5 = 18.5$.)

BACK
Median of ungrouped data · p. 441

RULE
C is the running total BEFORE the median class: never include its own frequency in C.

KEY
Locate the median class from the running total, then plug l, n, C, f, h into the formula.



In the median class 30 to 40, the formula's answer, 34.67, sits close to but not exactly at the class midpoint, 35.

CHECK YOURSELF

23. The median of grouped data is computed as

- **A.** $l + ((n - C)/f) \cdot h$, using n instead of $n/2$
- **B.** $l + ((n/2 - C)/f) \cdot h$, for the median class
- **C.** $((n/2 - C)/f) \cdot h$, without adding l
- **D.** $l + ((n/2 - C)/h) \cdot f$, with f and h swapped

24. 68 boxes are grouped 40-45, 45-50, 50-55, 55-60, 60-65 kg with frequencies 8, 14, 20, 15, 11. Here $n/2 = 34$, the median class is 50-55, with $C = 22$ (the total before it) and $f = 20$. The median is

- **A.** 61.5 kg, from using $n = 68$ instead of $n/2 = 34$
- **B.** 48 kg, from using the running total through 50-55 itself, $C = 42$, instead of the total before it
- **C.** 53 kg
- **D.** 50 kg, the lower limit of the median class, mistaken for the median

2 more in this set online.

Common mistakes

Any assumed mean gives the same mean

- **TRAP:** choosing a different value for the assumed mean, a, changes the mean.
- **REALITY:** it does not. Any valid choice of a gives exactly the same mean.

If we move a up by any amount, every deviation moves down by that same amount. The two changes cancel, so the mean stays the same. We pick a near the middle of the class marks only to keep the deviations small, never because one value is more correct than another.

Your turn: the homework table (frequencies $3, 5, 8, 3, 1$, marks 5 to 45) takes $a = 35$. What mean does it give? (Answer: 22, since the deviations are $-30, -20, -10, 0, 10$, their weighted total is -260, and $35 + (-260)/20 = 35 - 13 = 22$, the same 22 as every other route.)

DOES CHANGING THE ASSUMED MEAN CHANGE THE MEAN YOU COMPUTE?

WEAK

Two readers work the same table: classes 0–10, 10–20, 20–30, 30–40 with frequencies $2, 3, 4, 1$, so $n = 10$ and class marks $5, 15, 25, 35$. The first reader takes $a = 15$: the deviations are $-10, 0, 10, 20$, their weighted total is 40, and the mean is $15 + 40/10 = 19$. The second reader takes $a = 25$, but keeps the deviations already worked out from 15 instead of recalculating them, so gets $25 + 40/10 = 29$, and decides the choice of a changed the mean.

A correct working cannot depend on a, so two different means from two values of a mean one working is wrong. Moving a up by 10 must move every deviation down by 10.

STRONG

With $a = 25$ the deviations are $-20, -10, 0, 10$, their weighted total is -60, and the mean is $25 + (-60)/10 = 19$.

Both values of a give 19, the same as the direct method's $190/10 = 19$.

BACK

Mean of grouped data: the assumed-mean method · p. 444

KEY

Any assumed mean works: pick one that makes the deviations small, not a "correct" one.

A SCHOLAR INDIA REMEMBERS

Do you need the perfect assumed mean? Test it. Work the mean once with 30 as the assumed mean, and once with 40. You get the same mean both times. So the choice only changes how big the numbers are while you work. A class mark near the middle keeps them small.

Dignaga · the kingfisher

WATCH OUT

The trap: Getting $\overline{x} = 22$ with $a = 25$, the student expects a different mean when trying $a = 15$.

The correction: Any valid choice of a gives the same mean. With classes 0–10 to 40–50 and frequencies 3, 5, 8, 3, 1, $a = 25$ gives $\overline{x} = 22$ and $a = 15$ also gives $\overline{x} = 22$.

CHECK YOURSELF

25. A student computes the mean of homework-minutes data (classes 0-10 to 40-50, frequencies 3, 5, 8, 3, 1) twice by the assumed-mean method: once with $a = 25$, getting $\overline{x} = 22$, and once with $a = 15$. The student expects a different answer the second time, since a different a was chosen. What actually happens?

- **A.** the second calculation gives a different mean, since $a = 15$ is a smaller number than $a = 25$
- **B.** the second calculation also gives $\overline{x} = 22$ — any valid choice of a gives the same mean
- **C.** the second calculation cannot be completed, since $a = 15$ does not appear as a class mark in this table
- **D.** the second calculation gives $\overline{x} = 32$, since the smaller a needs 10 added back to match the first result

26. Why does changing the assumed mean a leave the final $\overline{x}$ unchanged, even though every deviation $d_i = x_i - a$ changes?

- **A.** because $\sum f_i$ also changes to compensate, keeping the ratio fixed
- **B.** because the class marks x_i change along with a
- **C.** the shift in d_i from a new a cancels when a is added back
- **D.** because only a very small range of values of a actually give the correct mean

1 more in this set online.

Read the frequencies either side of the modal class from the right rows

- **TRAP:** read f_0 off the modal class's own row, instead of the row immediately before it.
- **REALITY:** that gives a wrong mode. f_0 comes from the row just above the modal class, never from the modal row itself.

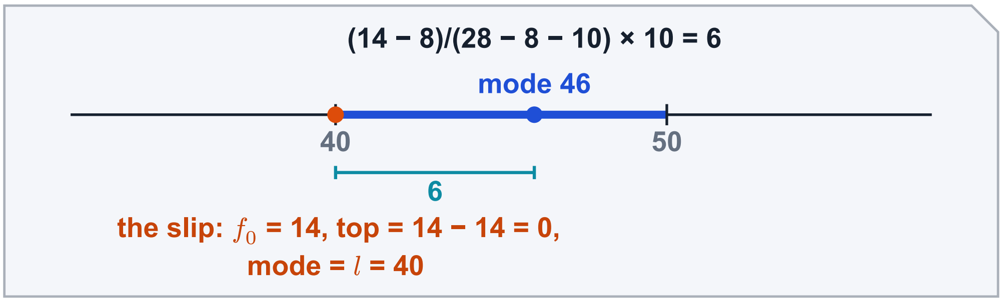
If we reuse the modal row's own frequency as f_0, the top of the fraction becomes zero. The mode then lands exactly on the lower limit l, not inside the modal class. Copying f_1 into f_0 or f_2 out of habit is the single most common slip in this formula.

Your turn: a modal class 40–50 has $f_1 = 14$, $f_0 = 8$ (the row before) and $f_2 = 10$ (the row after), with $h = 10$. What is the mode? (Answer: 46, since $40 + ((14 - 8)/(28 - 8 - 10)) \cdot 10 = 40 + 6 = 46$.)

CAUTION
Copying f_1 into f_0 or f_2 out of habit is the single most common mode-formula slip.

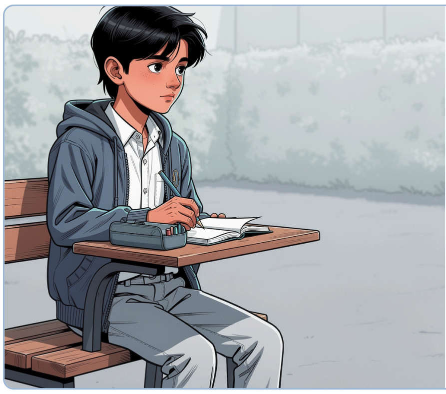
BACK
Mode of grouped data: the formula · p. 446

Reuse the modal row and the mode falls on the limit



With f_0 read from the row before, the mode sits 6 into the class at 46; copy f_1 into f_0 and the numerator is 14 − 14 = 0, so the mode drops onto the lower limit, 40.

Reusing the modal row's own frequency as fo drags the mode from 46 down onto the class's lower limit, 40.



FIRST TRY

~~Modal class 10 to 15, with $f_1 = 10$ and $f_2 = 5$. I read f_0 straight off the same row, $f_0 = 10$, and got mode $= 10 + (10 - 10)/(20 - 10 - 5) \cdot 5 = 10$.~~

SECOND LOOK

f_0 comes from the class just before the modal class, not the modal class itself. That class had frequency 6, so $f_0 = 6$, and mode $= 10 + (10 - 6)/(20 - 6 - 5) \cdot 5 = 10 + 4/9 \cdot 5 \approx 12.22$.

f_0 always comes from the row just above the modal class, never from the modal row.

WHICH TWO FREQUENCIES FEED THE MODE FORMULA, THE MODAL ROW OR ITS NEIGHBOURS?

WEAK

For classes 0–10, 10–20, 20–30, 30–40, 40–50 with frequencies $3, 7, 12, 8, 2$, the modal class is 20–30, since 12 is the highest frequency, so $l = 20$, $h = 10$, $f_1 = 12$. A reader takes $f_0 = 12$ as well, reading it from the modal row itself instead of the row before it. That makes the top of the fraction $12 - 12 = 0$, so the mode comes out as $20 + 0 = 20$, exactly on the lower limit.

The row before the modal class holds 7, not 12, so $f_0 = 12$ cannot be right. A mode landing exactly on the modal class's lower limit, with nothing pulling it inward, is a sign the wrong frequency went in.

STRONG

f_0 is the frequency of the class before the modal class, and f_2 is the frequency of the class after it. Here $f_0 = 7$ and $f_2 = 8$.

$20 + ((12 - 7)/(24 - 7 - 8)) \cdot 10 \approx 25.56$, a value inside the modal class, between 20 and 30, as it must be.

CHECK YOURSELF

27. For TV-watching hours grouped 0-5, 5-10, 10-15, 15-20, 20-25 with frequencies 2, 6, 10, 5, 2, the modal class is 10-15 with $f_1 = 10$. A student sets $f_0 = 10$, reading it off the modal class's own row, instead of 6, the frequency of the class immediately before it. What mode does each approach give?

- **A.** both methods give the same mode, 12.22, since f_0 cancels out of the formula either way
- **B.** the correct method gives mode $= 10$; the student's method gives mode $= 12.22$ instead
- **C.** the correct method gives mode $= 12.22$; the student's method gives mode $= 10$ instead
- **D.** the correct method gives mode $= 12.22$; the student's method gives mode $= 9$ instead

28. In the mode formula, why must f_0 come from the class immediately before the modal class, rather than any other class in the table?

- **A.** it compares the modal class against its two immediate neighbours only
- **B.** because f_0 must always be the smallest frequency in the entire table
- **C.** because the classes must be renumbered so the modal class becomes the first row
- **D.** because f_0 and f_1 must always be equal for the formula to apply

1 more in this set online.

Worked examples

WORKED EXAMPLE · Mean of homework time by the direct method

1.	classes 0–10, 10–20, 20–30, 30–40, 40–50 minutes; frequencies $3, 5, 8, 3, 1$; class marks $5, 15, 25, 35, 45$	*class marks read as the midpoint of each interval, from the frequency table for 20 students' homework time*
2.	$\sum f_i x_i = 3(5) + 5(15) + 8(25) + 3(35) + 1(45) = 15 + 75 + 200 + 105 + 45 = 440$	*multiply each class mark by its own frequency, then add every product*
3.	$\sum f_i = 20$	*add the frequency column to get the total number of students*
4.	$\overline{x} = 440/20 = 22$ minutes	*divide the total of $f_i x_i$ by the total frequency, the direct-method formula*
5.	check by the assumed-mean method with $a = 25$: $\sum f_i d_i = 3(-20) + 5(-10) + 8(0) + 3(10) + 1(20) = -60 - 50 + 0 + 30 + 20 = -60$	*Multiply each class mark's deviation from $a = 25$ by its own frequency, then add every product: the same homework-time table, a different route to the mean.*
6.	$\overline{x} = 25 + (-60/20) = 22$ minutes	CHECK ***Check the direct-method answer by a second route.*** *The assumed-mean method lands on the same 22 minutes, so the total added in step 2 was right.*

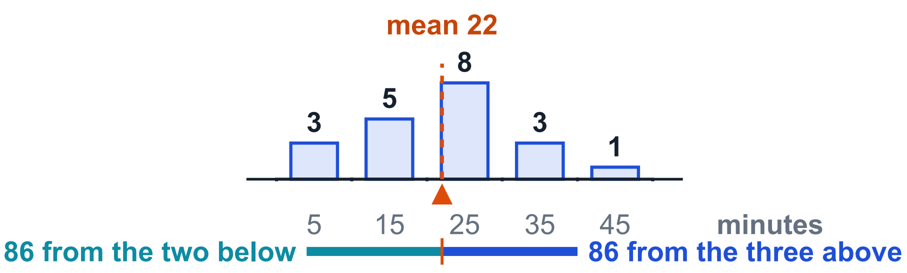
TRY
Classes 0–10, 10–20, 20–30 with frequencies 2, 3, 5. Find the mean, direct method.

Answer: $\sum f_i x_i = 180$, $\sum f_i = 10$, so $\overline{x} = 180/10 = 18$.

BACK
Mean of grouped data: the direct method · p. 443

RULE
List class mark and frequency side by side before multiplying: skipping this table is where slips get in.

The mean is the point where the table balances



Each class mark carries its own frequency as a weight. The three classes above 22 minutes pull with 86, the two below pull with 86, and the beam sits level only at 22. That balance is what $\sum fx$ divided by $\sum f$ finds, and it is the reason the direct method is worth running at all.

The mean is the one point on the ruler where the classes above and the classes below pull with equal force.

FIND THE MEAN OF GROUPED DATA BY THE DIRECT METHOD.

1. **List the class marks** Write down each class's class mark. For classes 0–10 to 40–50, the marks are $5, 15, 25, 35, 45$.

2. **Multiply by frequency** The frequencies are $3, 5, 8, 3, 1$. Multiply each class mark by its own frequency, giving $5(3) = 15$, $15(5) = 75$, $25(8) = 200$, $35(3) = 105$, $45(1) = 45$.

3. **Add both columns** Add the frequencies to get the total count, 20. Add the five products to get 440.

4. **Divide** Divide the total of the products by the total frequency. $440/20 = 22$ minutes.

5. **Check the size** The mean must sit among the class marks, closest to the tallest class. 22 sits between 15 and 25, near the class 20–30, the one with the highest frequency.

CHECK YOURSELF

29. 20 students' homework time is grouped 0-10, 10-20, 20-30, 30-40, 40-50 minutes with frequencies 3, 5, 8, 3, 1. By the direct method, $\sum f_i x_i =$

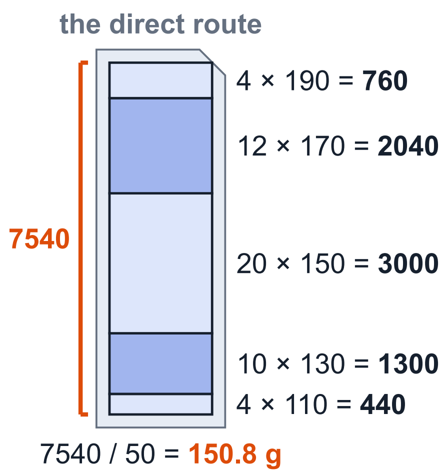
- **A.** 20, the value of $\sum f_i$, mistaken for $\sum f_i x_i$
- **B.** 125, the sum of the class marks alone, without weighting by frequency
- **C.** 22, the final mean, mistaken for the intermediate sum
- **D.** 440

30. In this worked example, why does the class 20-30 contribute the largest single term, $8(25) = 200$, to the running total $\sum f_i x_i$?

- **A.** 20-30 has the highest frequency, 8, so it is weighted most
- **B.** because 20-30 is the middle class in the table, and the middle class always contributes the most
- **C.** because 25 is the largest class mark among the five classes
- **D.** because the direct method always assigns extra weight to whichever class the assumed mean is chosen from

1 more in this set online.

The direct route stacks five products



Each block is one class's frequency times its class mark; the five stack to 7540 grams, and 7540 shared among 50 packets is 150.8 grams.

A stacked column of five weighted blocks totals 7540 grams, giving a mean of 150.8 grams over 50 packets.

WORKED EXAMPLE · **Mean tea-packet weight by the assumed-mean method**

1.	classes 100–120 to 180–200 grams; frequencies $4, 10, 20, 12, 4$; class marks $110, 130, 150, 170, 190$	*class marks read from the frequency table for 50 tea packets*
2.	take $a = 150$; deviations $d_i = -40, -20, 0, 20, 40$	*choose the middle class mark as the assumed mean, then subtract it from every class mark*
3.	$\sum f_i d_i = 4(-40) + 10(-20) + 20(0) + 12(20) + 4(40) = -160 - 200 + 0 + 240 + 160 = 40$	*multiply each deviation by its own frequency, then add every product*
4.	$\overline{x} = 150 + 40/50 = 150.8$ grams	*add the assumed mean to the total deviation divided by the total frequency*
5.	check by the direct method: $\sum f_i x_i = 4(110) + 10(130) + 20(150) + 12(170) + 4(190) = 7540$	*multiply each class mark by its own frequency directly, without any deviation shortcut*
6.	$7540/50 = 150.8$ grams	CHECK ***Check the assumed-mean answer against the direct method.*** *Both land on 150.8 grams, so the shortcut changed nothing.*

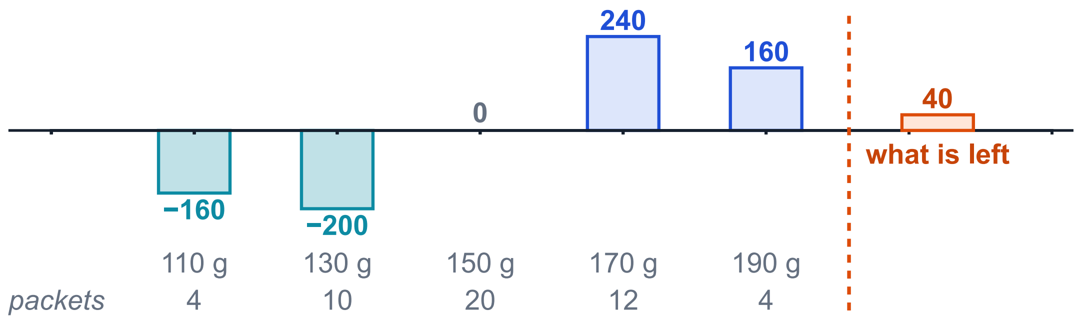
TRY
Classes 100–120, 120–140, 140–160 with frequencies 4, 10, 6. Mean, assumed-mean method, $a = 130$.

Answer: $d_i = -20, 0, 20$; $\sum f_i d_i = 40$, $\sum f_i = 20$, so $\overline{x} = 130 + 40/20 = 132$.

RULE
Choosing a near the middle of the class marks keeps the deviations small on both sides.

The deviations cancel, and only the leftover moves the mean

each class pulls by its own count times its own deviation from 150 g



The two classes below 150 g pull the total down by 360, the two above pull it up by 400, and almost all of that cancels. Only 40 survives, which over 50 packets is 0.8 g, so the mean is 150 + 0.8 = 150.8 g.

The two classes below the assumed mean nearly cancel the two above, and the small orange bar is what moves the answer off 150 grams.

FIND THE MEAN OF GROUPED DATA BY THE ASSUMED-MEAN METHOD.

1. **List the class marks** Write down each class's class mark. For the tea-packet classes, the marks are $110, 130, 150, 170, 190$.
2. **Choose the assumed mean** Pick a class mark near the middle as a. Here $a = 150$.

3. Find each deviation Subtract a from every class mark. $d_i = -40, -20, 0, 20, 40$.
4. Multiply and add The frequencies are $4, 10, 20, 12, 4$. Multiply each deviation by its own frequency and add the products. $4(-40) + 10(-20) + 20(0) + 12(20) + 4(40) = 40$.
5. Add back to a Add the frequencies to get 50. Divide the deviation total by 50 and add it to a. $150 + 40/50 = 150.8$ grams.
6. Check with a second route Work the same table with a different assumed mean, or by the direct method. Either route must give back the same 150.8 grams.

CHECK YOURSELF

31. 50 tea packets are grouped 100-120, 120-140, 140-160, 160-180, 180-200 g with frequencies 4, 10, 20, 12, 4. Taking $a = 150$, the deviations are $d_i = -40, -20, 0, 20, 40$, and $\sum f_i d_i = 40$. The mean weight is

- **A.** 150 g, the assumed mean with no correction added
- **B.** 0.8 g, the correction term alone, without adding back a
- **C.** 190 g, adding a and $\sum f_i d_i$ with the division by $\sum f_i$ skipped
- **D.** 150.8 g

32. Which single number in this worked example plays the role of a, the assumed mean, chosen for convenience rather than computed?

- **A.** 40, the frequency-weighted deviation total $\sum f_i d_i$
- **B.** 0.8, the correction added back to a
- **C.** 150.8, the final computed mean
- **D.** 150, the class mark of the middle class, 140-160

1 more in this set online.

WORKED EXAMPLE · **Mean delivery distance by the step-deviation method**

1.	classes 0–20 to 80–100 km; frequencies $6, 14, 22, 12, 6$; class marks $10, 30, 50, 70, 90$	*class marks read from the frequency table for 60 delivery riders*
2.	take $a = 50$ and $h = 20$; $u_i = -2, -1, 0, 1, 2$	*every deviation from $a = 50$ shares the common factor $h = 20$, so divide it out*
3.	$\sum f_i u_i = 6(-2) + 14(-1) + 22(0) + 12(1) + 6(2) = -12 - 14 + 0 + 12 + 12 = -2$	*multiply each u_i by its own frequency, then add every product*
4.	$\overline{x} = 50 + 20 \cdot (-2/60) \approx 49.33$ km	*the step-deviation formula, multiplying back by h before adding to a*
5.	check by the direct method: $\sum f_i x_i = 6(10) + 14(30) + 22(50) + 12(70) + 6(90) = 2960$	*multiply each class mark by its own frequency directly, without dividing out the class size*
6.	$2960/60 \approx 49.33$ km	CHECK ***Check the step-deviation answer against the direct method.*** *Both land on 49.33 km: it is a shortcut, not a different answer.*

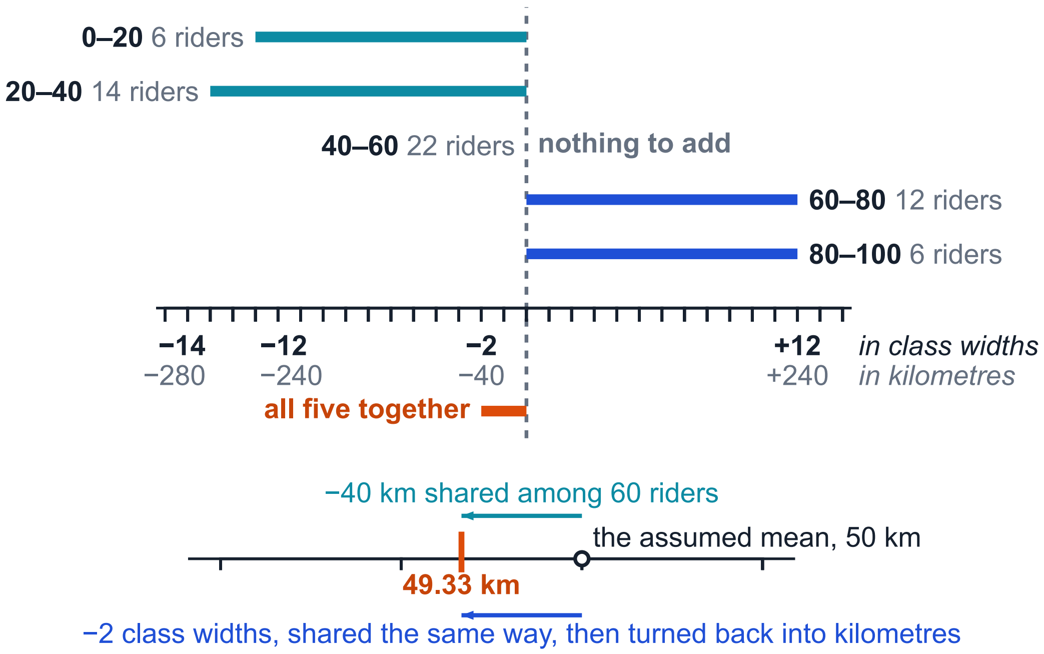
BACK
Mean of grouped data: the step-deviation method · p. 445

TRY
Classes 0–20, 20–40, 40–60 with frequencies 5, 10, 5. Mean, step-deviation method, $a = 30$.

Answer: $h = 20$, so $u_i = -1, 0, 1$; $\sum f_i u_i = 0$, so $\overline{x} = 30 + 20 \cdot (0/20) = 30$.

RULE
All five deviations here share the factor 20, the class size: exactly when this method pays off.

The numbers get smaller, the answer does not



Every bar is one class: how many riders, times how far its class mark sits from 50 km. Counted in kilometres the five leave −40 between them, and counted in class widths they leave −2, the same leftover with the 20 divided out. Share either one among 60 riders, put the 20 back where it was taken out, and the mean is 49.33 km both times.

The same five bars carry two rows of numbers, class widths and kilometres, both giving a mean of 49.33 km.

FIND THE MEAN OF GROUPED DATA BY THE STEP-DEVIATION METHOD.

1. **List the class marks** Write down each class's class mark. For the delivery classes, the marks are $10, 30, 50, 70, 90$.
2. **Choose a and read h** Pick a class mark near the middle as a, and read the class size h from the table. Here $a = 50$ and $h = 20$.
3. **Divide each deviation by h** Subtract a from each class mark, then divide by h. $u_i = -2, -1, 0, 1, 2$.
4. **Multiply and add** The frequencies are $6, 14, 22, 12, 6$. Multiply each u_i by its own frequency and add the products. $6(-2) + 14(-1) + 22(0) + 12(1) + 6(2) = -2$.
5. **Scale back up and add to a** Add the frequencies to get 60. Multiply the total from the step before by h, divide by 60, and add to a. $50 + 20 \cdot (-2/60) \approx 49.33$ km.
6. **Check by the direct method** Work the same table by the direct method. It must give back the same ≈ 49.33 km.

CHECK YOURSELF

33. 60 delivery riders' daily distance is grouped 0-20, 20-40, 40-60, 60-80, 80-100 km with frequencies 6, 14, 22, 12, 6. Taking $a = 50$, $h = 20$ gives $u_i = -2, -1, 0, 1, 2$ and $\sum f_i u_i = -2$. The mean distance is

- **A.** 49.33 km
- **B.** 50 km, the plain assumed mean value
- **C.** 49.97 km, from leaving out the multiplication by h
- **D.** 10 km, adding a to h times the raw total, division by $\sum f_i$ omitted

34. In this worked example, $\sum f_i u_i = -2$ comes out negative. What does that indicate about the mean, before finishing the calculation?

- **A.** the calculation has an error, since $\sum f_i u_i$ can never be negative
- **B.** the mean will land slightly below $a = 50$
- **C.** the mean will be far below a, by roughly $h = 20$ km
- **D.** the mean will be above $a = 50$, since negative values always increase a mean

1 more in this set online.

WORKED EXAMPLE · A missing frequency, found from a mean you already know

1.	classes 10–20 to 50–60 pots; frequencies $2, k, 13, 5, 2$; class marks $15, 25, 35, 45, 55$	*the clay pots fired in a day by a group of potters, with one frequency never recorded*
2.	$\sum f_i x_i = 2(15) + 13(35) + 5(45) + 2(55) + 25k = 820 + 25k$	*every class you can count goes in as a number, and the one you cannot keeps its letter*
3.	$\sum f_i = 2 + 13 + 5 + 2 + k = 22 + k$	*the total is short by the same unknown, so it carries the letter too*
4.	$820 + 25k = 34(22 + k)$	*the mean is given as 34 pots, and the mean times the total equals the sum of $f_i x_i$*
5.	$820 + 25k = 748 + 34k$, so $72 = 9k$ and $k = 8$	*multiply out the bracket, then gather the letters on one side and the numbers on the other*
6.	check by putting $k = 8$ back. $\sum f_i x_i = 820 + 200 = 1020$, $\sum f_i = 30$, and $1020/30 = 34$	CHECK ***Check the mean comes back as the 34 pots the question gave.*** *If it did not, the arithmetic went wrong, not the method.*

TRY

Classes 0–10, 10–20, 20–30 with frequencies $3, k, 5$, and the mean is 16. Find k.

Answer: $\sum f_i x_i = 140 + 15k$ and $\sum f_i = 8 + k$, so $140 + 15k = 16(8 + k)$, giving $k = 12$.

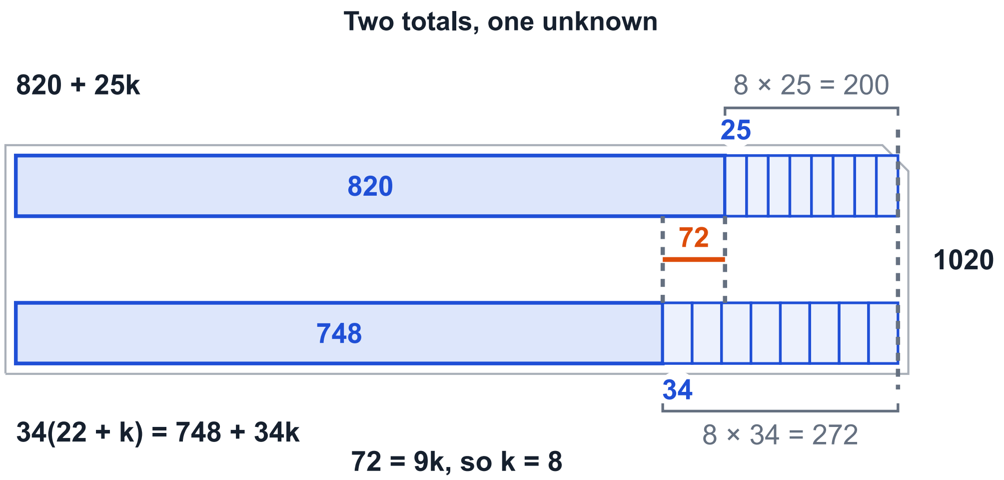
RULE

Write the unknown into both sums first. A missing frequency is the direct method with one letter left in.

Ananya counts the pots drying in a potter's courtyard, row by row.

FIND A MISSING FREQUENCY FROM A MEAN YOU ALREADY KNOW.

1. **Write the unknown into the table** Give the missing frequency a letter, k. For the pots table, the frequencies are $2, k, 13, 5, 2$ with class marks $15, 25, 35, 45, 55$.
2. **Total the known frequencies** Add the frequencies that are known numbers. $2 + 13 + 5 + 2 = 22$, so the total frequency is $22 + k$.
3. **Total the known products** Multiply the known frequencies by their class marks and add. $2(15) + 13(35) + 5(45) + 2(55) = 820$, plus $25k$ from the unknown class.
4. **Set mean times total equal to the sum** The mean is 34 pots. Write $820 + 25k = 34(22 + k)$.
5. **Solve for k** Multiply out the bracket and collect terms. $820 + 25k = 748 + 34k$, so $72 = 9k$ and $k = 8$.
6. **Check by putting k back** Put $k = 8$ back into the total frequency and the total of the products, then divide. The mean must come back as 34.



Both bars are the same total, 1020: the known products, 820, stand 72 ahead of 748, and each unknown pot adds 34 on the lower bar against 25 on the upper, so 8 of them close the 72.

Two equal-length bars for 820 plus 25k and 748 plus 34k close a gap of 72, giving k equals 8.

WORKED EXAMPLE · Mode of weekly TV-watching hours

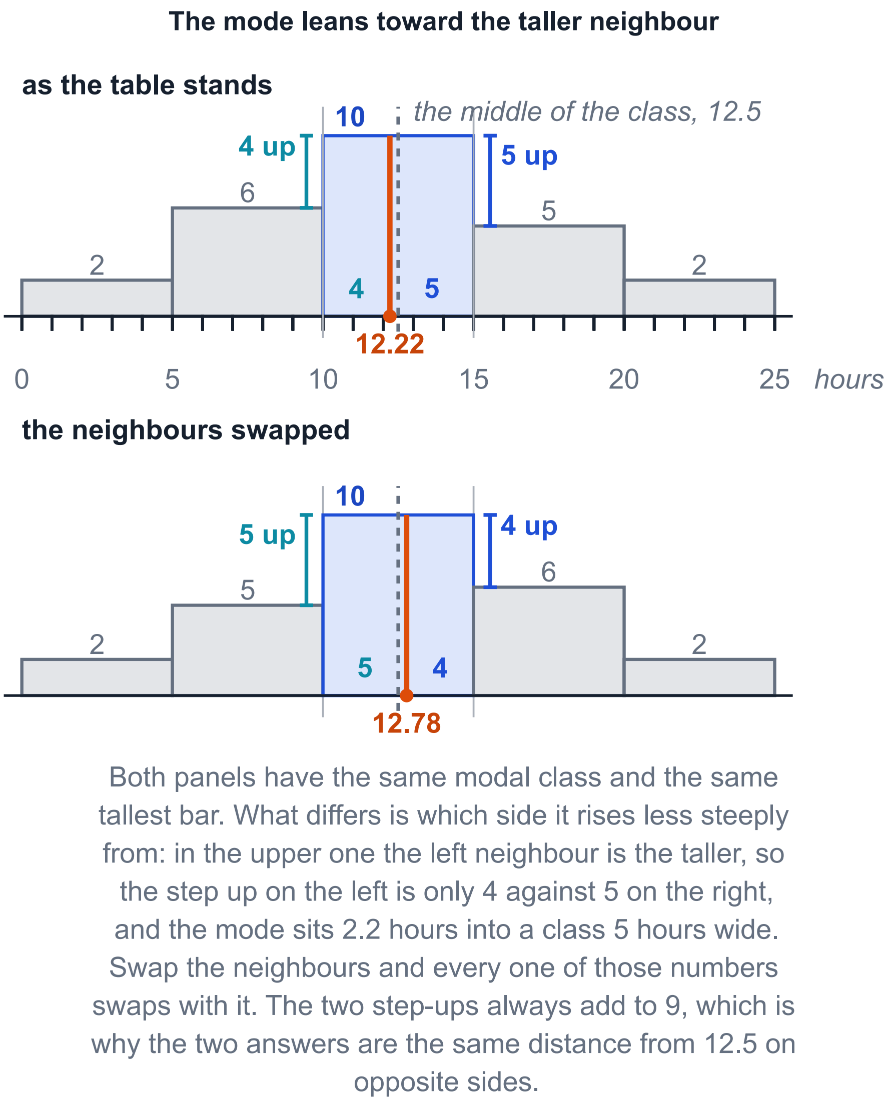
1.	classes 0–5 to 20–25 hours; frequencies $2, 6, 10, 5, 2$	*the frequency table for 25 people's weekly TV-watching hours*
2.	highest frequency 10 falls in the class 10–15 — the modal class	*scan the frequency column and pick its largest entry*
3.	$l = 10$, $h = 5$, $f_1 = 10$, $f_0 = 6$, $f_2 = 5$	*read the modal class's lower limit and class size, and its own and neighbouring frequencies, straight off the table*
4.	mode $= 10 + ((10 - 6)/(20 - 6 - 5)) \cdot 5 = 10 + (4/9) \cdot 5 \approx 12.22$ hours	*substitute into the mode formula for grouped data*
5.	$10 < 12.22 < 15$	CHECK ***Check the size of the answer.*** *The mode must sit inside the modal class it came from, never outside it, and here it does.*

BACK
Mode of grouped data: the formula · p. 446

TRY
Classes 0–5, 5–10, 10–15 with frequencies 3, 8, 4. Find the mode.

Answer: Modal class 5–10: $l = 5$, $f_1 = 8$, $f_0 = 3$, $f_2 = 4$, $h = 5$, mode $= 5 + (5/9) \cdot 5 \approx 7.78$.

RULE
Identify the modal class FIRST (highest frequency), then read f_0, f_2 off its immediate neighbours.



The two step-ups into the modal class split its base in their own ratio, so the mode leans toward the taller neighbour.

FIND THE MODE OF GROUPED DATA.

1. **Find the modal class** Scan the frequency column for the highest count. For the TV-watching table, frequencies $2, 6, 10, 5, 2$ peak at 10, in the class 10–15.
2. **Read l, h and f one** Read the modal class's lower limit, class size and own frequency. $l = 10$, $h = 5$, $f_1 = 10$.
3. **Read f nought and f two** Read the frequencies of the row before and the row after the modal class. $f_0 = 6$, $f_2 = 5$.
4. **Substitute** Put the five numbers into the mode formula. $10 + ((10 - 6)/(20 - 6 - 5)) \cdot 5 \approx 12.22$ hours.
5. **Check the size** The mode must sit inside the modal class, between 10 and 15. 12.22 does.

CHECK YOURSELF

35. 25 people's weekly TV hours are grouped 0-5, 5-10, 10-15, 15-20, 20-25 with frequencies 2, 6, 10, 5, 2. The modal class is 10-15, since

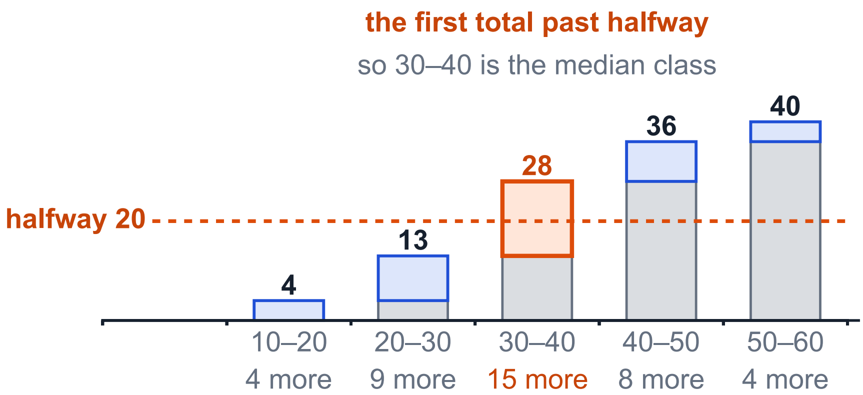
- **A.** its class mark, 12.5, is closest to the mean of all the class marks
- **B.** its frequency, 10, is the highest in the table
- **C.** it has the lowest frequency among its two neighbouring classes
- **D.** it is the only class with a frequency greater than 5

36. For the TV-hours data, the modal class 10-15 has $l = 10$, $h = 5$, $f_1 = 10$, $f_0 = 6$, $f_2 = 5$. The mode is

- **A.** 10 hours, from wrongly reusing $f_0 = 10$ instead of 6
- **B.** 12.78 hours, from swapping f_0 and f_2 in the formula
- **C.** 12.22 hours
- **D.** 12.5 hours, the class mark of the modal class, mistaken for the mode

1 more in this set online.

The running total is what names the median class



Before the class 30–40 only 13 members have been counted, which is short of halfway; once its own 15 are added the total is 28, which is past. That is the whole test, and it is what fixes l and C before the formula gives 34.67.

The median class is not read off the frequency column at all: it is the first class whose running total has got past halfway.

WORKED EXAMPLE · Median age of fitness-club members

1.	classes 10–20 to 50–60 years; frequencies $4, 9, 15, 8, 4$; $n = 40$	*the frequency table for 40 fitness-club members' ages*
2.	$n/2 = 20$; running total $4, 13, 28, 36, 40$	*add frequencies class by class to build the running total*
3.	running total first reaches 20 at 28, in the class 30–40 — the median class	*the first class where the running total is at least $n/2$*

WORKED EXAMPLE · **Median age of fitness-club members**

4.	$l = 30, h = 10, f = 15, C = 13$	*C is the running total through the class before the median class. f is the median class's own frequency.*
5.	median $= 30 + ((20 - 13)/15) \cdot 10 = 30 + (7/15) \cdot 10 \approx 34.67$ years	*substitute into the median formula for grouped data*
6.	$30 < 34.67 < 40$	CHECK ***Check the size of the answer.*** *The median must sit inside the median class it came from, never outside it, and here it does.*

BACK
Median of grouped data: the formula · p. 447

TRY
Classes 10–20, 20–30, 30–40 with frequencies 5, 10, 5. Find the median.

Answer: Median class 20–30: $l = 20$, $C = 5$, $f = 10$, $h = 10$, so median $= 20 + ((10 - 5)/10) \cdot 10 = 25$.

RULE
Compute $n/2$ first, then track the running total class by class until it first reaches that value.

FIND THE MEDIAN OF GROUPED DATA.

1. **Find n over two** Add all the frequencies to get n, then halve it. For the fitness-club table, $n = 40$, so $n/2 = 20$.
2. **Build the running total** Add the frequencies class by class. $4, 13, 28, 36, 40$.
3. **Find the median class** Find the first running total that reaches $n/2 = 20$. That is 28, in the class 30–40.
4. **Read l, h, f and C** Read the median class's lower limit and class size, its own frequency, and the running total before it. $l = 30$, $h = 10$, $f = 15$, $C = 13$.
5. **Substitute** Put the five numbers into the median formula. $30 + ((20 - 13)/15) \cdot 10 \approx 34.67$ years.
6. **Check the size** The median must sit inside the median class, between 30 and 40. 34.67 does.

CHECK YOURSELF

37. 40 fitness-club members' ages are grouped 10-20, 20-30, 30-40, 40-50, 50-60 with frequencies 4, 9, 15, 8, 4. Here $n/2 = 20$, and the median class is 30-40, since

- **A.** it is the class where the running total first EQUALS exactly 20
- **B.** the running total first passes 20 there, at 28
- **C.** it is the last class where the running total is below 20
- **D.** it is determined by comparing class marks to the mean, not by a running total

38. For the club-ages data, the median class 30-40 has $l = 30$, $h = 10$, $f = 15$, $C = 13$. The median age is

- **A.** 48 years, from using $n = 40$ instead of $n/2 = 20$
- **B.** 24.67 years, from using the running total through 30-40 itself, $C = 28$, instead of the total before it
- **C.** 34.67 years
- **D.** 35 years, the class mark of the median class, mistaken for the median

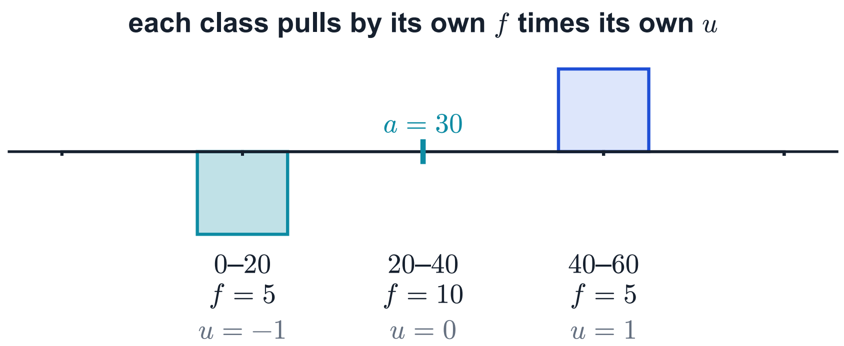
1 more in this set online.

Wrapping up

IN EVERY CHAPTER

Recap

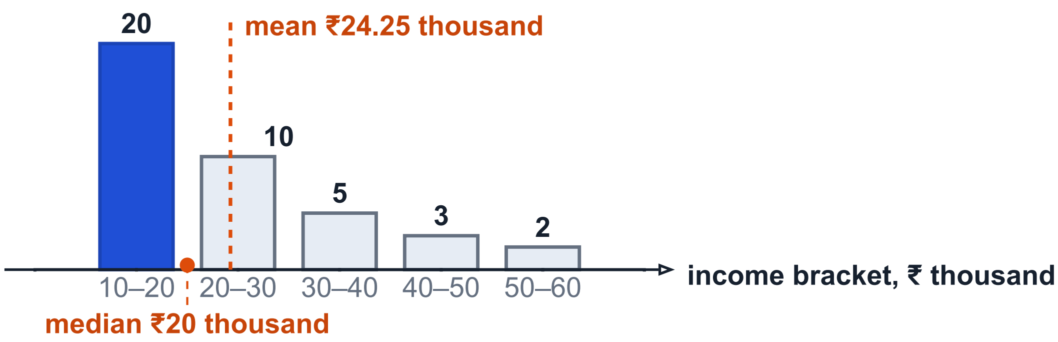
A symmetric table leaves nothing over



Frequencies $5, 10, 5$ sit symmetrically about $a = 30$, so $u = -1, 0, 1$ and $\sum fu = 0$. The mean is $30 + 20 \cdot (0/20) = 30$: a sits exactly on the mean.

Placing the assumed mean at the centre of a symmetric table makes the deviations cancel, and the correction term vanishes.

One skewed table, three measures, three different answers



This table is read as a discrete series, one point per income bracket, not through the interpolated grouped-median formula this chapter's median-class figure uses – the two are different tools for different figures, not two routes to the same number.

On this skewed income table, the mean (24.25 thousand rupees) and the median (20 thousand rupees) sit well apart.

- MEAN, DIRECT: $\overline{x} = (\sum f_i x_i)/(\sum f_i)$. For the homework-time table, $440/20 = 22$ minutes.

- MEAN, ASSUMED-MEAN OR STEP-DEVIATION: $\overline{x} = a + (\sum f_i d_i)/(\sum f_i)$. For the tea-packet table, $150 + 40/50 = 150.8$ grams.
- MODE: $l + ((f_1 - f_0)/(2f_1 - f_0 - f_2)) \cdot h$. For the TV-watching table, $10 + (4/9) \cdot 5 \approx 12.22$ hours.
- MEDIAN: $l + ((n/2 - C)/f) \cdot h$. For the fitness-club table, $30 + (7/15) \cdot 10 \approx 34.67$ years.
- CLASS MARK: the average of a class's lower and upper limits. For 10–25, $(10 + 25)/2 = 17.5$.

Every problem here starts from one grouped frequency table, and each formula reads off a different summary.

The mean multiplies each class mark by its frequency, directly or through the assumed-mean or step-deviation shortcut. Checking any one against another always gives agreement. The mode locates the modal class and its two neighbours. The median locates the median class from a running total, then applies one formula.

All three generalise the same three ideas we already met for ungrouped data: average, most frequent, and middle value. Grouping the data into class intervals only changes where the numbers come from, never what the three questions ask.

KEY
Same table, three questions: what's typical (the mean), what's most common (the mode), what's in the middle (the median).

BACK
Mean of grouped data: the direct method · p. 443

CHECK YOURSELF

39. This chapter's recap ties the mean, mode and median back to one shared object. What is it?

- **A.** three separate frequency tables, one for each measure
- **B.** the same list of raw, ungrouped values
- **C.** the same class mark for all three measures
- **D.** the same grouped frequency table

40. The mean has shortcut methods (assumed-mean, step-deviation) to ease hand computation. Why do the mode and median formulas need no equivalent shortcut?

- **A.** because the mode and median are always simpler ideas than the mean
- **B.** because the mode and median formulas do not use the class size h at all
- **C.** because a shortcut for the mode and median already exists but this chapter simply omits it
- **D.** they already use small numbers, from one or two classes

1 more in this set online.

CHECK YOURSELF: THE WHOLE CHAPTER

41. This chapter's methods for the mean, median and mode of grouped data all start from

- **A.** each raw value on its own, listed one by one, just as in ungrouped data
- **B.** the number of classes in the table, regardless of their frequencies
- **C.** a rough estimate, since grouped data can never give an exact value
- **D.** a frequency table of class intervals, not the raw values one by one

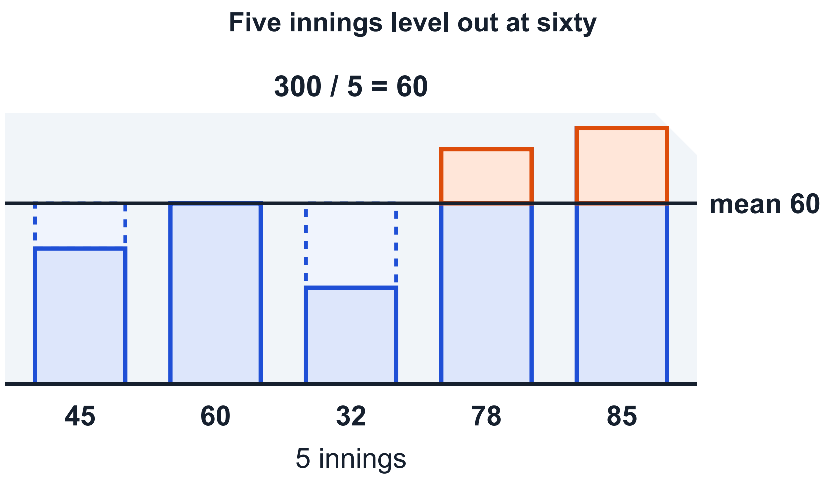
42. A frequency table has two classes with class marks 10 and 20, and frequencies 2 and 3. Using the direct method, what is the mean?

- **A.** $(10+20)/2=15$, the plain average of the two class marks, ignoring the frequencies entirely
- **B.** $(2\cdot 10+3\cdot 20)/(2+3)=16$, each class mark weighted by its own frequency
- **C.** $(2\cdot 10+3\cdot 20)/2=40$, the weighted sum divided by the number of classes instead of the total frequency
- **D.** $2\cdot 10+3\cdot 20=80$, the weighted sum, without dividing by anything

3 more in this set online.

IN EVERY CHAPTER

Where you will meet this



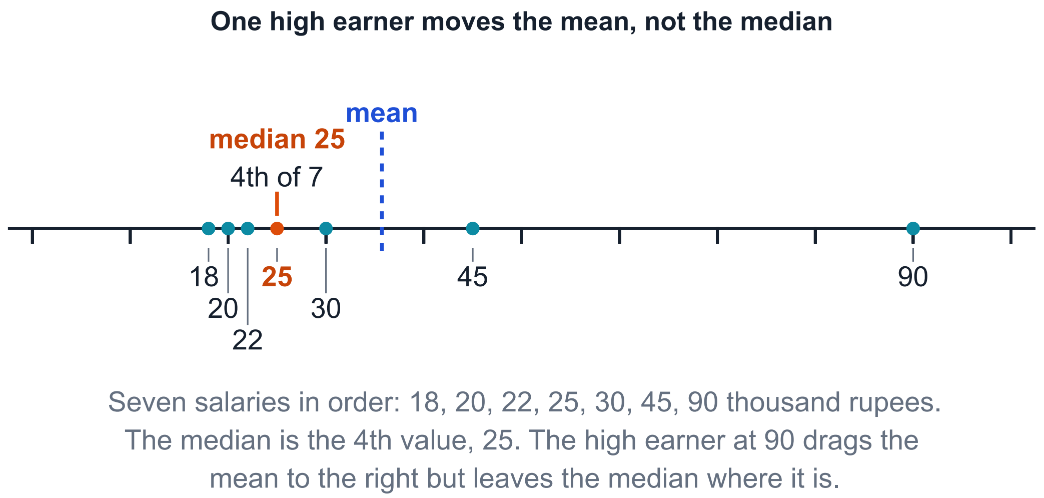
The two innings above 60 carry exactly what the two below it lack, so 300 runs shared over 5 innings is a level line at 60.

A batting average is the height the five scores would all have if the runs were shared out evenly, which is what the mean means.

You boil a list of numbers down to one figure more often than you think. Here are eight places a mean, a median or a mode gives you that number.

- **Working out a batting average.** A batter scores 45, 60, 32, 78 and 85 runs in 5 innings.
  The maths: a batting average is the mean, the sum of the scores over the number of innings.
  The mean is $(45+60+32+78+85)/5=300/5=60$ runs.
- **Finding the median salary at a small company.** A company's 7 employees earn, in thousands of rupees, $18, 20, 22, 25, 30, 45, 90$.
  The maths: the median is the middle salary once the list is in order. A high earner does not pull it up, unlike the mean.
  With 7 values in order, the median is the 4th value, 25 thousand rupees.

BACK
Mean of ungrouped data · p. 440



The median depends only on position, so one extreme value cannot move it; the mean depends on every value, so it can.

- **Finding a class's average test score.** A class of 30 students is grouped by marks. 2 scored 0–20, 8 scored 20–40, 12 scored 40–60, 6 scored 60–80 and 2 scored 80–100.
  The maths: the mean uses each class's mark, weighted by how many students fall in it.
  Using class marks $10, 30, 50, 70, 90$, the mean is $(2 \cdot 10 + 8 \cdot 30 + 12 \cdot 50 + 6 \cdot 70 + 2 \cdot 90)/30 \approx 48.67$.

- **Finding the best-selling price band in a shoe shop.** A shop sells shoes in five price bands. It sells 4 pairs in ₹1000–₹1200 and 10 in ₹1200–₹1400. It also sells 18 in ₹1400–₹1600, 9 in ₹1600–₹1800 and 5 in ₹1800–₹2000.
  The maths: the best-selling price sits in the modal class, and the formula narrows it to one number inside that band.
  With $l = 1400$, $h = 200$, $f_1 = 18$, $f_0 = 10$, $f_2 = 9$, the mode is $1400 + (8/17) \cdot 200 \approx 1494.12$.

- **Finding the median water bill in a colony.** A colony has 50 homes grouped by monthly water bill. 6 pay ₹400–₹600, 14 pay ₹600–₹800 and 16 pay ₹800–₹1000. 10 pay ₹1000–₹1200 and 4 pay ₹1200–₹1400.
  The maths: the median's value comes from the median class's limit, its running total and how many homes are in it.
  The running total passes 25 at ₹800–₹1000, so $800 + ((25 - 20)/16) \cdot 200 = 862.5$.

- **Finding the average rent in an apartment block.** Of 25 flats, monthly rent runs in thousands of rupees. 3 pay 15–20, 7 pay 20–25 and 10 pay 25–30. 4 pay 30–35 and 1 pays 35–40.
  The maths: when class marks are large, an assumed mean turns them into smaller deviations before averaging.
  Taking $a = 27.5$ thousand rupees, the deviations give $27.5 + (-35)/25 = 26.1$ thousand rupees.

- **Finding the average daily fare collected by a cab driver.** Over 50 days, fares fell in five bands. ₹200–₹300 on 5 days, ₹300–₹400 on 12 days and ₹400–₹500 on 20 days. ₹500–₹600 on 9 days and ₹600–₹700 on 4 days.
  The maths: dividing every deviation by the class size first keeps the numbers small to add.
  Taking $a = 450$ and $h = 100$, the mean is $450 + 100 \cdot (-5)/50 = 440$.

- **Setting a tiffin service's default roti count.** A tiffin service notes how many rotis each of 8 customers orders: $2, 3, 2, 4, 2, 3, 2, 1$.
  The maths: the count ordered most often is the mode, so it makes the best default to pack. 2 rotis appear 4 times, more than any other count, so the mode is 2.

Your turn. A shop's daily sales, in rupees, over 7 days are $2000, 1500, 1800, 2200, 1700, 2500, 1600$. What is the median daily sale? *Answer:* In order: ₹1500, ₹1600, ₹1700, ₹1800, ₹2000, ₹2200, ₹2500. The middle value is ₹1800.

Practice: Exercise 13.1

PRACTICE SET

1. PRACTICE The number of coconuts picked from each of 30 palms in an orchard is grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $2, 5, 10, 8, 5$. Find the mean number of coconuts per palm by the direct method. (Worked in full below — read it, then do the next two the same way.)

SOLUTION — Q1 · **Worked in full**

1.	$x_i = 5, 15, 25, 35, 45$	*the class mark is the midpoint of a class, halfway between its two limits, and it stands in for every reading in that class*
2.	$\sum f_i x_i = 2(5) + 5(15) + 10(25) + 8(35) + 5(45) = 10 + 75 + 250 + 280 + 225 = 840$	*multiply each class mark by its own frequency, then add every product*
3.	$\sum f_i = 2 + 5 + 10 + 8 + 5 = 30$	*add the frequency column, and check it comes to the 30 palms the question counted*
4.	$\overline{x} = 840/30 = 28$	*divide the total of $f_i x_i$ by the total frequency, which is the direct-method formula*
5.	the check — 28 lies between the smallest class mark 5 and the largest 45	*a mean outside that range is always wrong, and this one sits just past 25, the mark of the class holding the most palms*

2. PRACTICE The number of seats filled in each of 40 buses leaving a depot is grouped as 0–20, 20–40, 40–60, 60–80 and 80–100, with frequencies $5, 9, 12, 10, 4$. Find the mean number of seats filled by the assumed-mean method, taking $a = 50$. (The same three steps as above, with deviations from a in place of the class marks.)

6. PRACTICE The electricity bills (in ₹) of 40 households are grouped as 0–100, 100–200, 200–300, 300–400 and 400–500, with frequencies $6, 10, 14, 7, 3$. Find the mean bill by the step-deviation method, taking $a = 250$ and $h = 100$.

7. PRACTICE The daily screen time (in minutes) of a group of students is grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $5, k, 10, 7, 3$. If the mean screen time is 22 minutes, find k and the total number of students.

3. PRACTICE The water used in a day by each of 50 households (in litres) is grouped as 0–50, 50–100, 100–150, 150–200 and 200–250, with frequencies $7, 9, 16, 12, 6$. Find the mean water used by the step-deviation method, taking $a = 125$ and $h = 50$. (Divide each deviation by h before multiplying, and the last line puts the h back.)

4. PRACTICE The daily wages (in ₹) of 25 workers are grouped as 100–120, 120–140, 140–160, 160–180 and 180–200, with frequencies $4, 6, 8, 5, 2$. Find the mean daily wage by the direct method.

5. PRACTICE The marks obtained by 30 students in a test are grouped as 10–20, 20–30, 30–40, 40–50 and 50–60, with frequencies $3, 7, 12, 5, 3$. Find the mean marks by the assumed-mean method, taking $a = 35$.

8. PRACTICE The marks scored by 20 students in an assignment are grouped as 0–5, 5–10, 10–15, 15–20 and 20–25, with frequencies $2, 4, 8, 4, 2$. Find the mean marks by the direct method.

9. PRACTICE The ages (in years) of 45 employees at a factory are grouped as 20–25, 25–30, 30–35, 35–40 and 40–45, with frequencies $6, 12, 15, 8, 4$. Find the mean age by the step-deviation method, taking $a = 32.5$ and $h = 5$.

ANSWERS

1. Class marks $5, 15, 25, 35, 45$. $\sum f_i x_i = 840$ and $\sum f_i = 30$, so the mean is $840/30 = 28$ coconuts — between the smallest class mark and the largest, as it must be.

2. Deviations $-40, -20, 0, 20, 40$ about $a = 50$. $\sum f_i d_i = -20$ and $\sum f_i = 40$, so the mean is $50 + (-20/40) = 49.5$ seats.

3. Step deviations $-2, -1, 0, 1, 2$ about $a = 125$ with $h = 50$. $\sum f_i u_i = 1$ and $\sum f_i = 50$, so the mean is $125 + (1/50)(50) = 126$ litres.

4. Class marks $110, 130, 150, 170, 190$. $\sum f_i x_i = 3650$, $\sum f_i = 25$. Mean $= 3650/25 = 146$ rupees.

5. Class marks $15, 25, 35, 45, 55$; deviations $-20, -10, 0, 10, 20$ from $a = 35$. $\sum f_i d_i = -20$, $\sum f_i = 30$. Mean $= 35 + (-20/30) = 34.33$.

6. Class marks $50, 150, 250, 350, 450$; $u_i = -2, -1, 0, 1, 2$. $\sum f_i u_i = -9$, $\sum f_i = 40$. Mean $= 250 + 100 \cdot (-9/40) = 227.5$ rupees.

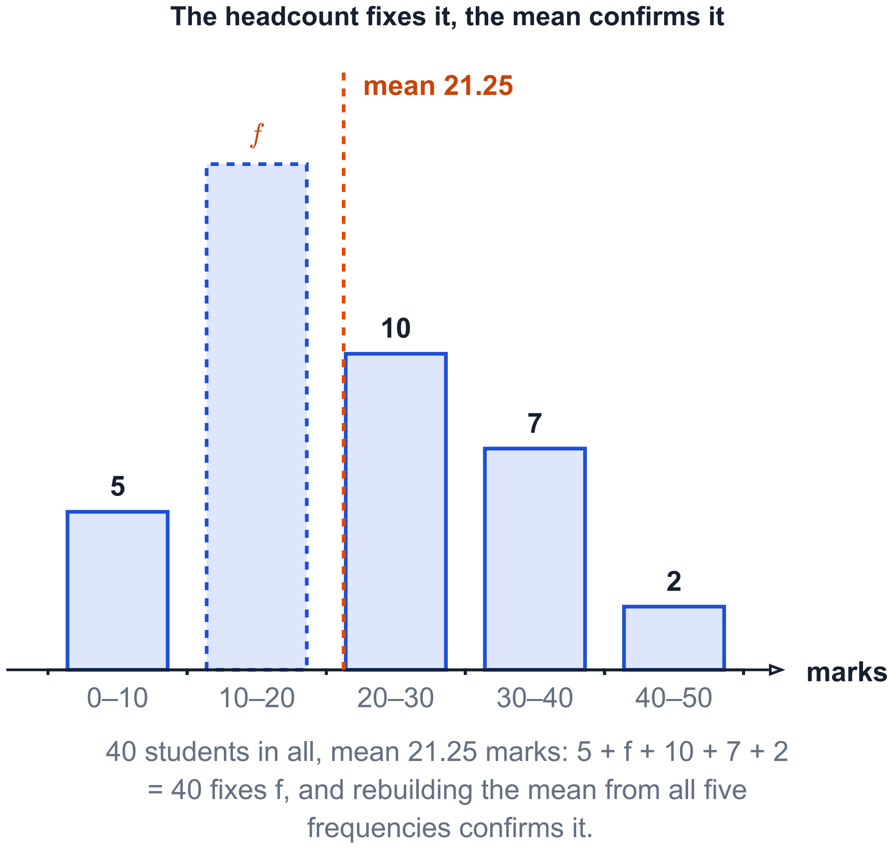
7. The mean equation gives $655 + 15k = 22(25 + k)$, so $7k = 105$ and $k = 15$. Total students $= 25 + k = 40$.

8. Class marks $2.5, 7.5, 12.5, 17.5, 22.5$. $\sum f_i x_i = 250$, $\sum f_i = 20$. Mean $= 250/20 = 12.5$.

9. Class marks $22.5, 27.5, 32.5, 37.5, 42.5$; $u_i = -2, -1, 0, 1, 2$ from $a = 32.5$. $\sum f_i u_i = -8$, $\sum f_i = 45$. Mean $= 32.5 + 5 \cdot (-8/45) \approx 31.61$.

Full solutions: **manthika.com/learn/math10-ch13**

Further practice



Four bars are solid and one is drawn as a dashed outline, its height still unknown, with the mean marked between two of the classes.

PRACTICE SET

10. PRACTICE The number of books read in a year by 20 students of a class are grouped as 0–4, 4–8, 8–12, 12–16 and 16–20 books, with frequencies $3, 5, 6, 4, 2$. Find the mean number of books read, by the direct method.

SOLUTION — Q10 · Worked in full

1.	$\sum f_i x_i = 3(2) + 5(6) + 6(10) + 4(14) + 2(18) = 6 + 30 + 60 + 56 + 36 = 188$	*Multiply each class mark by its own frequency, then add every product.*
2.	$\sum f_i = 3 + 5 + 6 + 4 + 2 = 20$	*Add the frequency column to get the total number of students.*
3.	$\overline{x} = 188/20 = 9.4$	*Divide the total of $f_i x_i$ by the total frequency, the direct-method formula.*

11. PRACTICE The maximum daily temperature (in degrees Celsius) recorded over 40 days is grouped as 15–20, 20–25, 25–30, 30–35 and 35–40, with frequencies $5, 8, 15, 8, 4$. Find the mean temperature by the assumed-mean method, taking $a = 27.5$.

15. PRACTICE The speeds (in kilometres per hour) of 40 vehicles crossing a toll plaza are grouped as 40–50, 50–60, 60–70, 70–80 and 80–90, with frequencies $4, 9, 15, 8, 4$. Find the mean speed by the step-deviation method, taking $a = 65$ and $h = 10$.

12. PRACTICE The number of minutes 30 patients waited at a clinic is grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $4, 7, 10, 6, 3$. Find the mean waiting time by the step-deviation method, taking $a = 25$ and $h = 10$.

13. PRACTICE The number of customers visiting a shop each day over 20 days is grouped as 10–20, 20–30, 30–40 and 40–50, with frequencies $5, 8, 4, 3$. What is the mean number of customers, by the direct method?

A. 25
B. 27.5
C. 30
D. 32.5

14. PRACTICE The daily rainfall (in millimetres) recorded over 30 days in a town is grouped as 0–8, 8–16, 16–24, 24–32 and 32–40, with frequencies $3, 6, 10, 7, 4$. Find the mean rainfall by the assumed-mean method, taking $a = 20$.

16. PRACTICE The monthly savings (in ₹) of 50 employees are grouped as 1000–1500, 1500–2000, 2000–2500, 2500–3000 and 3000–3500, with frequencies $6, 12, 17, 10, 5$. Find the mean saving by the direct method, then check it by the step-deviation method, taking $a = 2250$ and $h = 500$.

17. PRACTICE The number of minutes spent on homework each day by a group of students is grouped as 20–30, 30–40, 40–50, 50–60 and 60–70, with frequencies $8, 12, k, 10, 6$. If the mean homework time is 43.8 minutes, find k.

18. PRACTICE The number of minutes spent exercising daily by 75 members of a gym is grouped as 0–10, 10–20, 20–30, 30–40, 40–50 and 50–60, with frequencies $8, 12, p, 20, q, 5$. If the mean exercise time is 31 minutes, find p and q.

ANSWERS

10. Class marks $2, 6, 10, 14, 18$. $\sum f_i x_i = 188$, $\sum f_i = 20$. Mean $= 188/20 = 9.4$ books.

11. Class marks $17.5, 22.5, 27.5, 32.5, 37.5$; deviations $-10, -5, 0, 5, 10$ from $a = 27.5$. $\sum f_i d_i = -10$, $\sum f_i = 40$. Mean $= 27.5 + (-10/40) = 27.25$ degrees Celsius.

12. Class marks $5, 15, 25, 35, 45$; $u_i = -2, -1, 0, 1, 2$. $\sum f_i u_i = -3$, $\sum f_i = 30$. Mean $= 25 + 10 \cdot (-3/30) = 24$ minutes.

13. B — 27.5.

14. Class marks $4, 12, 20, 28, 36$; deviations $-16, -8, 0, 8, 16$ from $a = 20$. $\sum f_i d_i = 24$, $\sum f_i = 30$. Mean $= 20 + 24/30 = 20.8$ millimetres.

15. Class marks $45, 55, 65, 75, 85$; $u_i = -2, -1, 0, 1, 2$. $\sum f_i u_i = -1$, $\sum f_i = 40$. Mean $= 65 + 10 \cdot (-1/40) = 64.75$ kilometres per hour.

16. Class marks $1250, 1750, 2250, 2750, 3250$. Direct method: $\sum f_i x_i = 110500$, $\sum f_i = 50$, mean $= 110500/50 = 2210$. Step-deviation check: $u_i = -2, -1, 0, 1, 2$, $\sum f_i u_i = -4$, mean $= 2250 + 500 \cdot (-4/50) = 2210$ rupees — the two methods agree.

17. Class marks $25, 35, 45, 55, 65$. $\sum f_i x_i = 1560 + 45k$, $\sum f_i = 36 + k$. The mean equation $(1560 + 45k)/(36 + k) = 43.8$ gives $1.2k = 16.8$, so $k = 14$.

18. The frequencies give $p + q = 30$. The mean equation $(1195 + 25p + 45q)/75 = 31$ gives $25p + 45q = 1130$; substituting $p = 30 - q$ gives $q = 19$ and $p = 11$.

Full solutions: **manthika.com/learn/math10-ch13**

Practice: Exercise 13.2

PRACTICE SET

1. PRACTICE The vegetables sold in a day at a stall (in kilograms), recorded over 50 days, are grouped as 10–20, 20–30, 30–40, 40–50 and 50–60, with frequencies $5, 12, 20, 8, 5$. Find the modal weight sold. (Worked in full below — read it, then do the next two the same way.)

SOLUTION — Q1 · **Worked in full**

1.	the biggest frequency is 20, in the class 30–40, so 30–40 is the modal class	*the mode lives in the class with the most readings, so find that class before touching the formula*
2.	$l = 30$, $h = 10$, $f_1 = 20$, $f_0 = 12$, $f_2 = 8$	*l and h are the modal class's lower limit and width, f_1 is its own frequency, f_0 the class before it and f_2 the class after it*
3.	mode $= 30 + ((20 - 12)/(2(20) - 12 - 8)) \cdot 10 = 30 + (8/20) \cdot 10 = 34$	*substitute into the grouped-mode formula, and reduce the fraction before multiplying by h*
4.	the check — 34 lies inside the modal class 30–40	*the mode always lands inside its own class, and below the middle here because the class before it is busier than the class after it*

2. PRACTICE The number of trips made in a day by each of 40 auto-rickshaw drivers is grouped as 5–10, 10–15, 15–20, 20–25 and 25–30, with frequencies $4, 10, 18, 6, 2$. Find the modal number of trips. (Find the modal class first, then read its two neighbours off the table.)
3. PRACTICE The distance walked in a day by each of 50 postal workers (in kilometres) is grouped as 0–4, 4–8, 8–12, 12–16 and 16–20, with frequencies $8, 20, 12, 6, 4$. Find the modal distance walked. (The modal class is not the middle one here — read the frequency column before you decide.)
4. PRACTICE The number of goals scored by 30 football teams in a season is grouped as 0–2, 2–4, 4–6, 6–8 and 8–10, with frequencies $6, 10, 8, 4, 2$. Find the modal number of goals, and also the mean number of goals, for comparison.
5. PRACTICE The weekly allowance (in ₹) of 50 children is grouped as 20–40, 40–60, 60–80, 80–100 and 100–120, with frequencies $8, 14, 16, 7, 5$. Find the modal weekly allowance.
6. PRACTICE The number of runs scored by 35 batters in a tournament is grouped as 0–20, 20–40, 40–60, 60–80 and 80–100, with frequencies $5, 9, 12, 6, 3$. Find the modal number of runs.
7. PRACTICE The electricity units consumed by 40 households in a month are grouped as 0–50, 50–100, 100–150, 150–200 and 200–250 units, with frequencies $6, 9, 14, 7, 4$. Find the modal electricity consumption.
8. PRACTICE The number of plants in 30 gardens of a colony is grouped as 0–3, 3–6, 6–9, 9–12 and 12–15, with frequencies $4, 9, 10, 5, 2$. Find the modal number of plants, and also the mean, for comparison.

ANSWERS

1. Modal class 30–40, with $l = 30$, $h = 10$, $f_1 = 20$, $f_0 = 12$ and $f_2 = 8$. Mode $= 30 + (8/20) \cdot 10 = 34$ kilograms, which lies inside 30–40.
2. Modal class 15–20, with $l = 15$, $h = 5$, $f_1 = 18$, $f_0 = 10$ and $f_2 = 6$. Mode $= 15 + (8/20) \cdot 5 = 17$ trips.
3. Modal class 4–8, the second class and not the middle one, with $l = 4$, $h = 4$, $f_1 = 20$, $f_0 = 8$ and $f_2 = 12$. Mode $= 4 + (12/20) \cdot 4 = 6.4$ kilometres.
4. Modal class 2–4, $f_1 = 10$, $f_0 = 6$, $f_2 = 8$, $l = 2$, $h = 2$. Mode $= 2 + ((10 - 6)/(20 - 6 - 8)) \cdot 2 = 3.33$. Mean: class marks $1, 3, 5, 7, 9$, $\sum f_i x_i = 122$, mean $= 122/30 = 4.07$.
5. Modal class 60–80, $f_1 = 16$, $f_0 = 14$, $f_2 = 7$, $l = 60$, $h = 20$. Mode $= 60 + ((16 - 14)/(32 - 14 - 7)) \cdot 20 = 63.64$ rupees.
6. Modal class 40–60, $f_1 = 12$, $f_0 = 9$, $f_2 = 6$, $l = 40$, $h = 20$. Mode $= 40 + ((12 - 9)/(24 - 9 - 6)) \cdot 20 = 46.67$ runs.
7. Modal class 100–150, $f_1 = 14$, $f_0 = 9$, $f_2 = 7$, $l = 100$, $h = 50$. Mode $= 100 + ((14 - 9)/(28 - 9 - 7)) \cdot 50 = 120.83$ units.
8. Modal class 6–9, $f_1 = 10$, $f_0 = 9$, $f_2 = 5$, $l = 6$, $h = 3$. Mode $= 6 + ((10 - 9)/(20 - 9 - 5)) \cdot 3 = 6.5$. Mean: class marks $1.5, 4.5, 7.5, 10.5, 13.5$, $\sum f_i x_i = 201$, mean $= 201/30 = 6.7$.

Full solutions: **manthika.com/learn/math10-ch13**

BACK
Mode of grouped data: the formula · p. 446

Further practice

PRACTICE SET

9. PRACTICE The number of minutes spent on household chores daily by 50 teenagers is grouped as 0–5, 5–10, 10–15, 15–20 and 20–25, with frequencies $6, 12, 20, 8, 4$. Find the modal time spent on chores.

SOLUTION — Q9 · **Worked in full**

1.	$10 + ((20 - 12)/(40 - 12 - 8)) \cdot 5 = 10 + (8/20) \cdot 5 = 12$	*Substitute the modal class's lower limit $l = 10$, class size $h = 5$, its own frequency $f_1 = 20$ and its neighbours' frequencies $f_0 = 12$, $f_2 = 8$.*

10. PRACTICE The number of late arrivals recorded over 30 days at a train station is grouped as 0–5, 5–10, 10–15, 15–20 and 20–25, with frequencies $4, 9, 11, 4, 2$. Which class is the modal class?

A. 0–5
B. 5–10
C. 10–15
D. 15–20

13. PRACTICE The number of text messages sent daily by 60 teenagers is grouped as 0–20, 20–40, 40–60, 60–80 and 80–100, with frequencies $8, 15, 22, 10, 5$. Find the modal number of messages sent, correct to two decimal places.

11. PRACTICE The number of books borrowed per day from a library over 18 days is grouped as 0–6, 6–12, 12–18, 18–24 and 24–30 books, with frequencies $2, 3, 9, 3, 1$. Find the modal number of books borrowed.

12. PRACTICE The daily temperature (in degrees Celsius) recorded over 40 days in a hill town is grouped as 20–30, 30–40, 40–50, 50–60 and 60–70, with frequencies $5, 8, 16, 8, 3$. Find the modal temperature, and also the mean temperature, for comparison.

14. PRACTICE The number of minutes spent on social media daily by 45 students is grouped as 0–8, 8–16, 16–24, 24–32 and 32–40, with frequencies $8, 10, 18, 6, 3$. Find the modal time spent, and also the mean time, correct to two decimal places, for comparison.

15. PRACTICE For a delivery-time distribution, the modal class is 20–30 minutes, with lower limit $l = 20$, class size $h = 10$, modal-class frequency $f_1 = 25$, and the frequency of the class after it, $f_2 = 15$. If the mode works out to 25 minutes, find f_0, the frequency of the class before the modal class.

ANSWERS

9. Modal class 10–15, $f_1 = 20$, $f_0 = 12$, $f_2 = 8$, $l = 10$, $h = 5$. Mode $= 10 + ((20 - 12)/(40 - 12 - 8)) \cdot 5 = 12$ minutes.

10. C — 10–15.

11. Modal class 12–18, $f_1 = 9$, $f_0 = 3$, $f_2 = 3$, $l = 12$, $h = 6$. Mode $= 12 + ((9 - 3)/(18 - 3 - 3)) \cdot 6 = 15$ books.

12. Modal class 40–50, $f_1 = 16$, $f_0 = 8$, $f_2 = 8$, $l = 40$, $h = 10$. Mode $= 40 + ((16 - 8)/(32 - 8 - 8)) \cdot 10 = 45$ degrees Celsius. Mean: class marks $25, 35, 45, 55, 65$, $\sum f_i x_i = 1760$, mean $= 1760/40 = 44$ degrees Celsius.

13. Modal class 40–60, $f_1 = 22$, $f_0 = 15$, $f_2 = 10$, $l = 40$, $h = 20$. Mode $= 40 + ((22 - 15)/(44 - 15 - 10)) \cdot 20 \approx 47.37$ messages.

14. Modal class 16–24, $f_1 = 18$, $f_0 = 10$, $f_2 = 6$, $l = 16$, $h = 8$. Mode $= 16 + ((18 - 10)/(36 - 10 - 6)) \cdot 8 = 19.2$ minutes. Mean: class marks $4, 12, 20, 28, 36$, $\sum f_i x_i = 788$, mean $= 788/45 \approx 17.51$ minutes.

15. From the mode formula, $25 = 20 + ((25 - f_0)/(35 - f_0)) \cdot 10$, which gives $f_0 = 15$.

Full solutions: **manthika.com/learn/math10-ch13**

Practice: Exercise 13.3

PRACTICE SET

1. PRACTICE The heights of 50 saplings in a nursery (in centimetres) are grouped as 10–20, 20–30, 30–40, 40–50 and 50–60, with frequencies $7, 8, 20, 10, 5$. Find the median height. (Worked in full below — read it, then do the next two the same way.)

SOLUTION — Q1 · **Worked in full**

1.	$n = 7 + 8 + 20 + 10 + 5 = 50$, so $n/2 = 25$	*the median splits the readings into two halves, so the half-way count comes first*
2.	running totals $7, 15, 35, 45, 50$, and the first past 25 is 35, at the class 30–40	*add the frequencies downwards, and the class where the running total first reaches $n/2$ is the median class*
3.	$l = 30$, $h = 10$, $f = 20$, $C = 15$	*C is the running total BEFORE the median class, not counting the median class itself, which is the commonest place to slip*
4.	median $= 30 + ((25 - 15)/20) \cdot 10 = 30 + (10/20) \cdot 10 = 35$	*substitute into the grouped-median formula*
5.	the check — 35 lies inside the median class 30–40	*the running total is 15 before that class, short of 25, and 35 once it is counted, past 25, so the class was read correctly*

2. PRACTICE The number of tomatoes picked from each of 60 plants is grouped as 0–20, 20–40, 40–60, 60–80 and 80–100, with frequencies $10, 12, 16, 14, 8$. Find the median number of tomatoes. (Running totals first, then the class where they pass $n/2$.)
3. PRACTICE The masses of 40 guavas from an orchard (in grams) are grouped as 100–120, 120–140, 140–160, 160–180 and 180–200, with frequencies $5, 7, 16, 8, 4$. Find the median mass. (Take C from the class before the median class, never from the median class itself.)
4. PRACTICE The weights (in kg) of 40 wrestlers are grouped as 40–45, 45–50, 50–55, 55–60 and 60–65, with frequencies $4, 10, 14, 8, 4$. Find the median weight.
5. PRACTICE The daily sales (in ₹ thousand) of 50 shops are grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $6, 11, 18, 10, 5$. Find the median daily sales.
6. PRACTICE The number of hours of sleep reported by 45 students is grouped as 4–5, 5–6, 6–7, 7–8 and 8–9 hours, with frequencies $3, 8, 20, 10, 4$. Find the median sleep time.
7. PRACTICE The monthly electricity bills (in ₹) of 60 households are grouped as 100–150, 150–200, 200–250, 250–300 and 300–350, with frequencies $8, 14, 20, 12, 6$. Find the median bill, and also the mean and the modal bill, comparing all three.
8. PRACTICE The daily temperature (in °C) recorded over 35 days is grouped as 20–24, 24–28, 28–32, 32–36 and 36–40, with frequencies $4, 8, 13, 7, 3$. Find the median temperature.

ANSWERS

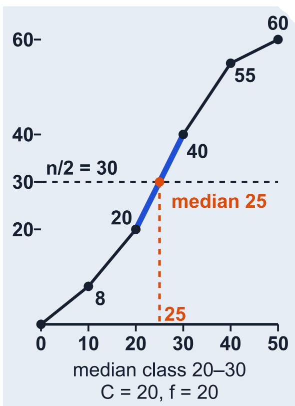
1. Running totals $7, 15, 35, 45, 50$ and $n/2 = 25$, so the median class is 30–40, with $l = 30$, $h = 10$, $f = 20$ and $C = 15$. Median $= 30 + (10/20) \cdot 10 = 35$ centimetres.
2. Running totals $10, 22, 38, 52, 60$ and $n/2 = 30$, so the median class is 40–60, with $l = 40$, $h = 20$, $f = 16$ and $C = 22$. Median $= 40 + (8/16) \cdot 20 = 50$ tomatoes.
3. Running totals $5, 12, 28, 36, 40$ and $n/2 = 20$, so the median class is 140–160, with $l = 140$, $h = 20$, $f = 16$ and $C = 12$. Median $= 140 + (8/16) \cdot 20 = 150$ grams.
4. $n = 40$, $n/2 = 20$. Running total $4, 14, 28, 36, 40$ first reaches 20 at 28, class 50–55. $l = 50$, $h = 5$, $f = 14$, $C = 14$. Median $= 50 + ((20 - 14)/14) \cdot 5 = 52.14$ kg.
5. $n = 50$, $n/2 = 25$. Running total $6, 17, 35, 45, 50$ first reaches 25 at 35, class 20–30. $l = 20$, $h = 10$, $f = 18$, $C = 17$. Median $= 20 + ((25 - 17)/18) \cdot 10 = 24.44$.
6. $n = 45$, $n/2 = 22.5$. Running total $3, 11, 31, 41, 45$ first reaches 22.5 at 31, class 6–7. $l = 6$, $h = 1$, $f = 20$, $C = 11$. Median $= 6 + ((22.5 - 11)/20) \cdot 1 = 6.58$ hours.
7. $n = 60$, $n/2 = 30$. Running total $8, 22, 42, 54, 60$ first reaches 30 at 42, class 200–250. $l = 200$, $h = 50$, $f = 20$, $C = 22$. Median $= 200 + ((30 - 22)/20) \cdot 50 = 220$. Mean: class marks $125, 175, 225, 275, 325$, $\sum f_i x_i = 13200$, mean $= 13200/60 = 220$. Mode: modal class 200–250, $f_1 = 20$, $f_0 = 14$, $f_2 = 12$, mode $= 200 + ((20 - 14)/(40 - 14 - 12)) \cdot 50 = 221.43$.
8. $n = 35$, $n/2 = 17.5$. Running total $4, 12, 25, 32, 35$ first reaches 17.5 at 25, class 28–32. $l = 28$, $h = 4$, $f = 13$, $C = 12$. Median $= 28 + ((17.5 - 12)/13) \cdot 4 = 29.69$ °C.

Full solutions: **manthika.com/learn/math10-ch13**

BACK
Median of grouped data: the formula · p. 447

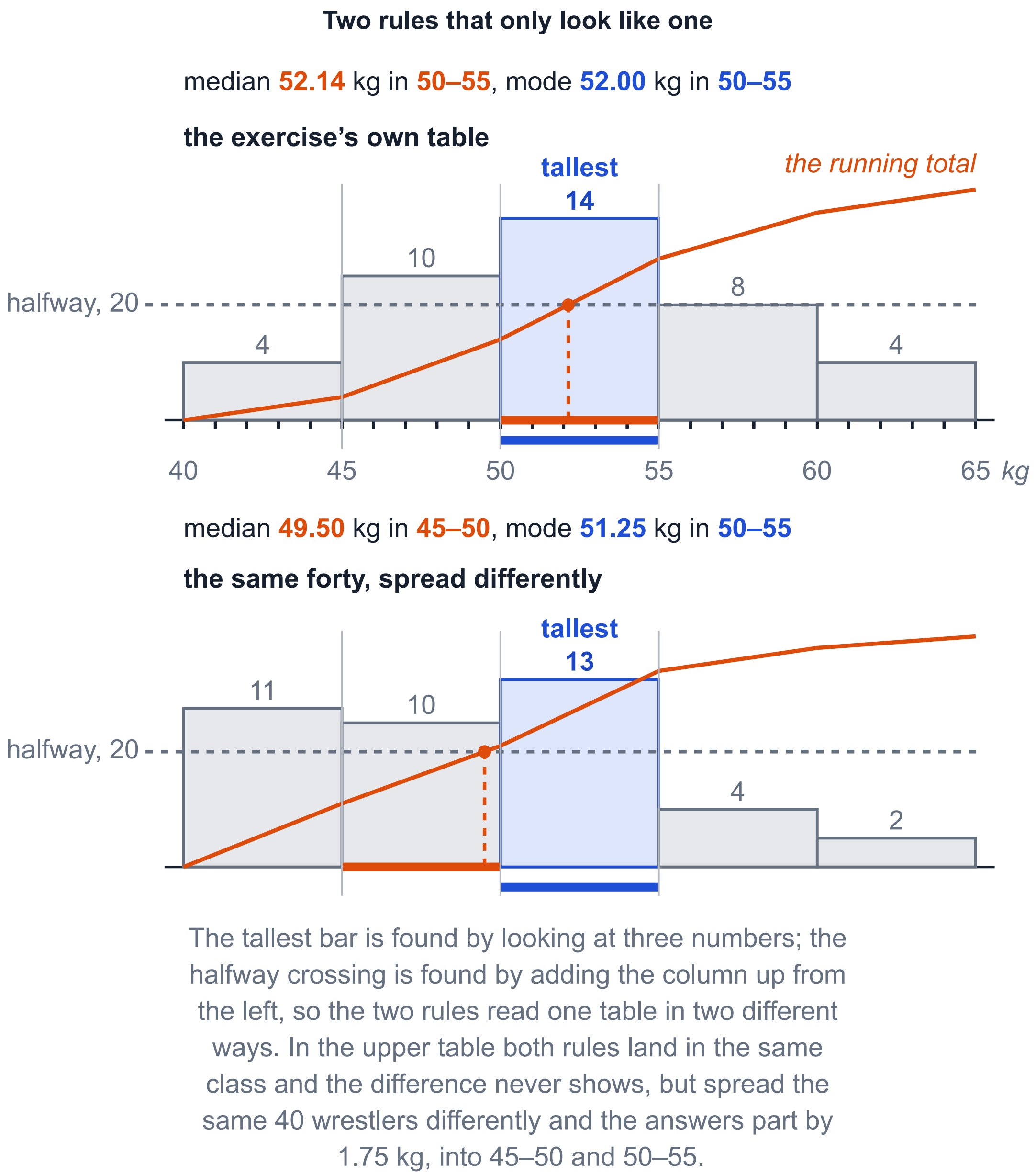
Further practice

The running total crosses thirty in the class twenty to thirty



The running totals 8, 20, 40, 55, 60 climb through n/2 = 30 between 20 and 40, so the median class is 20–30 with C = 20, and the formula lands at 25.

The median formula reads a straight line between two running totals, so 25 sits where the climb from 20 to 40 crosses half of 60.



Two similar frequency tables give a mode and a median that sometimes land on the same class and sometimes land on different ones.

PRACTICE SET

9. PRACTICE The number of minutes spent commuting daily by 60 office workers is grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $8, 12, 20, 15, 5$. Find the median commuting time.

SOLUTION — Q9 · **Worked in full**

1. $20 + ((30 - 20)/20) \cdot 10 = 20 + (10/20) \cdot 10 = 25$

Substitute the median class's lower limit $l = 20$, class size $h = 10$, its own frequency $f = 20$, and the cumulative frequency $C = 20$ before it, with $n/2 = 30$.

10. PRACTICE The daily screen time (in minutes) of 40 primary-school children is grouped as 0–8, 8–16, 16–24, 24–32 and 32–40, with frequencies $5, 9, 16, 7, 3$. Find the median screen time.

11. PRACTICE The number of visitors to a museum each day over 50 days is grouped as 0–20, 20–40, 40–60, 60–80 and 80–100, with frequencies $6, 10, 18, 11, 5$. Which class is the median class?

A. 0–20
B. 20–40
C. 40–60
D. 60–80

12. PRACTICE The number of pages read daily by 35 readers in a book club is grouped as 0–6, 6–12, 12–18, 18–24 and 24–30 pages, with frequencies $4, 7, 12, 9, 3$. Find the median number of pages read.

13. PRACTICE The waiting time (in minutes) of 50 patients at a clinic is grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $6, 9, 20, 10, 5$. Find the median waiting time, and also the mean waiting time, for comparison.

14. PRACTICE The number of minutes spent stretching before practice, recorded for 50 athletes, is grouped as 0–5, 5–10, 10–15, 15–20 and 20–25, with frequencies $5, 10, 18, 12, 5$. Find the median stretching time and the modal stretching time, both correct to two decimal places, for comparison.

15. PRACTICE The number of minutes spent reading daily by a group of students is grouped as 0–10, 10–20, 20–30, 30–40 and 40–50, with frequencies $5, 10, 20, 12, x$. If the median reading time is 25 minutes, find x.

16. PRACTICE The weights (in kilograms) of 50 parcels handled by a courier office are grouped as 0–10, 10–20, 20–30, 30–40, 40–50 and 50–60, with frequencies $2, 5, p, 12, q, 3$. If the median weight is 40 kilograms, find p and q.

ANSWERS

9. $n = 60$, $n/2 = 30$. Running total $8, 20, 40, 55, 60$ first reaches 30 at 40, class 20–30. $l = 20$, $h = 10$, $f = 20$, $C = 20$. Median $= 20 + ((30 - 20)/20) \cdot 10 = 25$ minutes.

10. $n = 40$, $n/2 = 20$. Running total $5, 14, 30, 37, 40$ first reaches 20 at 30, class 16–24. $l = 16$, $h = 8$, $f = 16$, $C = 14$. Median $= 16 + ((20 - 14)/16) \cdot 8 = 19$ minutes.

11. C — 40–60.

12. $n = 35$, $n/2 = 17.5$. Running total $4, 11, 23, 32, 35$ first reaches 17.5 at 23, class 12–18. $l = 12$, $h = 6$, $f = 12$, $C = 11$. Median $= 12 + ((17.5 - 11)/12) \cdot 6 = 15.25$ pages.

13. $n = 50$, $n/2 = 25$. Running total $6, 15, 35, 45, 50$ first reaches 25 at 35, class 20–30. $l = 20$, $h = 10$, $f = 20$, $C = 15$. Median $= 20 + ((25 - 15)/20) \cdot 10 = 25$ minutes. Mean: class marks $5, 15, 25, 35, 45$, $\sum f_i x_i = 1240$, mean $= 1240/50 = 24.8$ minutes.

14. $n = 50$, $n/2 = 25$. Running total $5, 15, 33, 45, 50$ first reaches 25 at 33, class 10–15. $l = 10$, $h = 5$, $f = 18$, $C = 15$. Median $= 10 + ((25 - 15)/18) \cdot 5 \approx 12.78$ minutes. Mode: same modal class, $f_1 = 18$, $f_0 = 10$, $f_2 = 12$. Mode $= 10 + ((18 - 10)/(36 - 10 - 12)) \cdot 5 \approx 12.86$ minutes.

15. The median class is 20–30 with $l = 20$, $h = 10$, $f = 20$, $C = 15$ (fixed, since x sits in the class after it). The median equation gives $n/2 = 25$, so $n = 50$ and $x = 50 - 47 = 3$.

16. The frequencies give $p + q = 28$. The median class is 30–40 with $l = 30$, $h = 10$, $f = 12$, $C = 7 + p$. The median equation gives $p = 6$, then $q = 28 - 6 = 22$.

Full solutions: **manthika.com/learn/math10-ch13**