IGCSE MATHEMATICS 0580 · CHAPTER 12 NOTES · ~34 MIN READ

Statistics

Averages, frequency tables, charts and scatter graphs, then the Extended topics: quartiles, cumulative frequency, box plots and histograms. Every method has a diagram and a worked example, so you can see what to do before you try it.

Averages & range | Frequency tables | Charts | Scatter diagrams | Cumulative frequency | Histograms

RS by Riyaz Sir
mentorriyaz.in

No tag = **Core** (everyone) EXTENDED Extended only highlight = comes up again and again

Test yourself at the end

IN 30 SECONDS

- **Mean** = total ÷ number of values. **Median** = middle value of the **ordered** data. **Mode** = most common. **Range** = largest − smallest. The median is the safest average when there are outliers.
- **Tables and charts:** for a frequency table use **f × x** to find the mean, and for grouped data use **midpoints** (the answer is an **estimate**). A pie chart angle is **frequency ÷ total × 360°**.
- **Extended: IQR = Q3 − Q1**. On a cumulative frequency graph read at **¼n, ½n and ¾n**. In a histogram the **area** of a bar is the frequency, so the height is **frequency density = frequency ÷ class width**. EXT

BY THE END YOU CAN

- Calculate the mean, median, mode and range from a list, a frequency table and a grouped table.
- Choose the most suitable average and explain why.
- Draw and read bar charts, pie charts and stem-and-leaf diagrams.
- Describe correlation, draw a line of best fit and use it to make predictions.
- Find quartiles and the interquartile range from a list. EXT
- Draw and use cumulative frequency graphs and box plots to estimate and compare. EXT
- Draw and read histograms with unequal class widths. EXT

1 Averages and spread

Types of data

Categorical data — described in **words or categories**, such as eye colour or favourite sport.

Discrete data — **numbers you count**, so only certain values are possible, such as the number of siblings or shoe size.

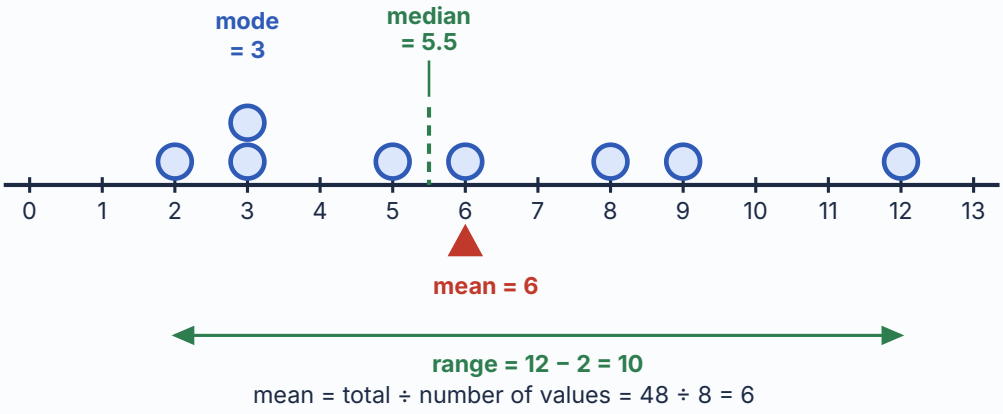
Continuous data — **numbers you measure**, which can take any value in a range, such as height, mass or time.

Type	How you get it	Examples
Categorical	sort into named groups	eye colour, type of transport
Discrete	count	goals scored, number of pets
Continuous	measure	height, time, mass

- **Grouped tables use continuous classes** for measured data, written like $10 < x \leq 20$. A value of exactly 10 goes in the class **before**.
- **A small sample is not reliable.** If a question asks you to comment on a conclusion from a few results, say that the **sample is too small** to be sure.

Mean, median, mode and range

Data: 2, 3, 3, 5, 6, 8, 9, 12



mean = total ÷ number of values = 48 ÷ 8 = 6

Sort the data first. The median is the middle position, the mode is the tallest pile and the mean is the balance point.

Mean — the total of all the values **divided by how many** values there are.

Median — the **middle value** when the data is in order. For an even number of values, take the **mean of the two middle values**.

Mode — the value that occurs **most often**. There can be two modes, or none at all.

Range — **largest value − smallest value**. It measures spread, and it is **not** an average.

$$\text{mean} = \frac{\text{sum of all the values}}{\text{number of values}}$$

add everything up, then divide

$$\text{median position} = \frac{n + 1}{2}$$

n is the number of values, and the data must be in order

- **Always put the data in order before finding the median.** This is the most common way to lose marks.

- **Check your mean.** It must lie between the smallest and largest values.
- **Reverse problems:** total = mean × number of values. Use it to find a missing value or to combine groups.

WORKED EXAMPLE

Find the mean, median, mode and range of 7, 3, 9, 3, 12, 5, 8, 3.

1. total = 7 + 3 + 9 + 3 + 12 + 5 + 8 + 3 = 50, so mean = 50 ÷ 8 = **6.25**
2. in order: 3, 3, 3, 5, 7, 8, 9, 12
3. median = (5 + 7) ÷ 2 = **6**
4. mode = **3**
5. **range = 12 − 3 = 9**

WORKED EXAMPLE

The mean of five numbers is 12. Four of them are 8, 10, 15 and 11. Find the fifth.

1. total of all five = 5 × 12 = 60
2. total of the four = 8 + 10 + 15 + 11 = 44
3. **fifth number = 60 − 44 = 16**

WORKED EXAMPLE

8 students have a mean mark of 62 and 12 students have a mean mark of 72. Find the mean of all 20 students.

1. total of the first group = 8 × 62 = 496
2. total of the second group = 12 × 72 = 864
3. total = 1360
4. **mean = 1360 ÷ 20 = 68**

WORKED EXAMPLE

Aisha has a mean of 70 after 4 tests. What must she score in the fifth test to have a mean of 72 after 5 tests?

1. total needed = 5 × 72 = 360
2. total so far = 4 × 70 = 280
3. **fifth test = 360 − 280 = 80**

COMMON MISTAKE

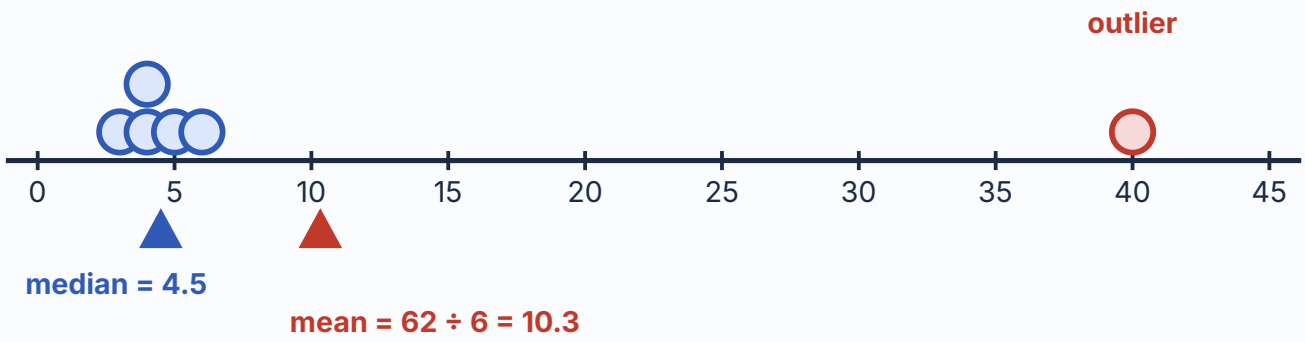
Averaging the averages. In the third example the answer is **not** (62 + 72) ÷ 2 = 67, because the two groups are different sizes.

EXAM TIP

For the median of an even number of values, write down the two middle values, then give their mean. For example, the middle values 5 and 7 give 6.

Choosing the best average

Data: 3, 4, 4, 5, 6, 40



One extreme value drags the mean up. The median hardly moves.

One extreme value, called an outlier, drags the mean along with it. The median stays in the middle of the group.

Outlier — a value that is **much larger or smaller** than the rest of the data.

Average	Best used when	Weakness
Mean	the data has no extreme values	pulled towards outliers
Median	there are outliers or the data is skewed	does not use every value
Mode	you want the most popular choice, or the data is categorical	there may be no mode, or more than one

- **The mean uses every value,** so one extreme value changes it a lot. The median only depends on the middle.
- **Use the mode for "most popular"** or "most common", such as the shoe size a shop should stock most of.
- **Give a reason in context.** "The median is better because the salary of 95 000 is an outlier" gets the mark. "It is better" does not.

WORKED EXAMPLE

The salaries at a small firm are 18 000, 20 000, 21 000, 22 000, 24 000 and 95 000. Find the mean and the median. Which gives a better idea of a typical salary?

1. total = 18 000 + 20 000 + 21 000 + 22 000 + 24 000 + 95 000 = 200 000
2. mean = 200 000 ÷ 6 = **33 333** to the nearest whole number
3. median = (21 000 + 22 000) ÷ 2 = **21 500**
4. **the median is better because 95 000 is an outlier, and five of the six people earn much less than the mean**

PICTURE IT

Imagine the data on a see-saw. The **mean** is the balance point, so a heavy outlier far from the middle pulls it a long way. The **median** only cares about the order, so it does not move.

2 Frequency tables

Ungrouped frequency tables

Quiz scores of 30 students

score x	f	f × x	cumulative f
1	4	4	4
2	7	14	11
3	10	30	21
4	6	24	27
5	3	15	30
total	**30**	**87**	

The orange row has the largest f, so it is the mode.

Mean

Σf × x ÷ Σf
= 87 ÷ 30
= 2.9

Median

(30 + 1) ÷ 2 = 15.5th
15th and 16th values are both in the row where cumulative f reaches 21, so median = 3

Mode = 3, range = 5 − 1 = 4

Add an f × x column to find the total, and a cumulative frequency column to find the median.

$$\text{mean} = \frac{\Sigma(f \times x)}{\Sigma f}$$

Σ means "the sum of". Σf is the total frequency

$$\text{median position} = \frac{\Sigma f + 1}{2}$$

then use the cumulative frequency column to find which row it is in

- **Mode** = the value with the **largest frequency**, not the largest frequency itself. In the diagram the mode is 3, not 10.
- **Range** = largest value − smallest value. Use the x column.
- **f × x** gives the total of each row. Add the column to get the total of all the data.

WORKED EXAMPLE

The number of goals in 20 matches is shown. 0 goals: 3 matches, 1 goal: 6, 2 goals: 7, 3 goals: 3, 4 goals: 1. Find the mean, median, mode and range.

1. Σf = 3 + 6 + 7 + 3 + 1 = 20
2. Σ(f × x) = 0 + 6 + 14 + 9 + 4 = 33, so mean = 33 ÷ 20 = **1.65**
3. cumulative frequency: 3, 9, 16, 19, 20
4. the 10th and 11th values are both in the row for 2 goals, so median = **2**
5. mode = **2** (largest frequency 7)
6. **range = 4 − 0 = 4**

WORKED EXAMPLE

A frequency table has x = 1, 2, 3, 4 with frequencies 5, k, 6 and 4. The mean is 2.5. Find k.

1. Σf = 5 + k + 6 + 4 = 15 + k
2. Σ(f × x) = 5 + 2k + 18 + 16 = 39 + 2k
3. mean: (39 + 2k) ÷ (15 + k) = 2.5
4. 39 + 2k = 37.5 + 2.5k, so 1.5 = 0.5k
5. **k = 3**
6. **check: Σf = 18, Σ(f × x) = 45 and 45 ÷ 18 = 2.5**

COMMON MISTAKE

Using the **frequency** as the mode. The mode is the **score** with the highest frequency.

COMMON MISTAKE

Dividing by the **number of rows** instead of by Σf. The mean is Σ(f × x) ÷ Σf.

Grouped frequency tables

Time to complete a puzzle (seconds)

time x	f	midpoint	f × midpoint
0 < x ≤ 10	4	5	20
10 < x ≤ 20	9	15	135
20 < x ≤ 30	15	25	375
30 < x ≤ 40	8	35	280
40 < x ≤ 50	4	45	180
total	**40**		**990**

Estimated mean

Σ f × midpoint ÷ Σf
= 990 ÷ 40
= 24.75

Modal class

largest frequency:
20 < x ≤ 30

Median class

20th and 21st values:
cumulative f is 4, 13, 28
so 20 < x ≤ 30

Use the midpoint of each class as a stand-in for all the values in it. The result is an estimate.

Modal class — the class with the **largest frequency**. It is not a single value.

Median class — the class that **contains the median**. Find it with the cumulative frequency.

$$\text{estimated mean} = \frac{\Sigma(f \times \text{midpoint})}{\Sigma f}$$

add a midpoint column and an f × midpoint column

$$\text{midpoint} = \frac{\text{lower boundary} + \text{upper boundary}}{2}$$

the middle of the class

- **It is only an estimate** because you do not know the exact values. You assume they are all at the midpoint.
- **For the median class** with 40 values, the median is halfway between the 20th and 21st values.
- **Check your answer.** The estimated mean must lie inside the range of the classes, usually near the modal class.

WORKED EXAMPLE

The heights of 50 plants are grouped. 0 < h ≤ 10: 6 plants, 10 < h ≤ 20: 14, 20 < h ≤ 30: 18, 30 < h ≤ 40: 12. Estimate the mean, and give the modal class and the median class.

1. midpoints are 5, 15, 25 and 35
2. Σ(f × midpoint) = 30 + 210 + 450 + 420 = 1110
3. estimated mean = 1110 ÷ 50 = **22.2**
4. modal class = **20 < h ≤ 30** (frequency 18)
5. **25th and 26th values: cumulative frequency is 6, 20, 38, so the median class is 20 < h ≤ 30**

COMMON MISTAKE

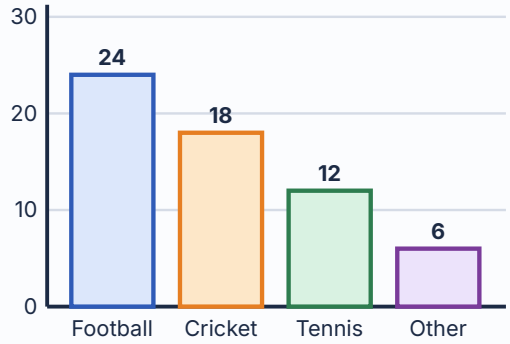
Using the **class width** or the **upper boundary** as the midpoint. Add the two boundaries and halve the answer.

Riyaz Sir

3 Charts and diagrams

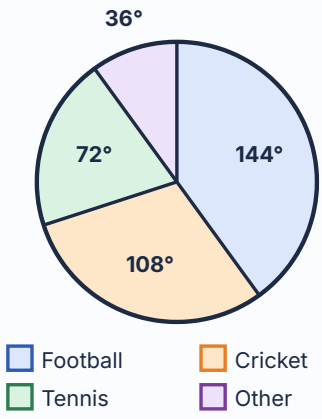
Bar charts and pie charts

Bar chart



bar height = frequency

Pie chart



angle = frequency ÷ total × 360°

A bar chart shows the frequency as a height. A pie chart shows each share as an angle out of 360°.

$$\text{angle} = \frac{\text{frequency}}{\text{total frequency}} \times 360°$$

for a pie chart

$$\text{frequency} = \frac{\text{angle}}{360°} \times \text{total frequency}$$

to read a pie chart

- **Bar charts:** equal-width bars with equal gaps, a scale that starts at **0**, and labels on both axes.
- **Pictograms:** each symbol stands for a number of items, so always read the **key**. A half symbol is half of that number.
- **Pie charts** show proportions. You can only find real frequencies if you know the total, or one sector and its angle.
- **Check:** all the angles in a pie chart must add up to **360°**.

WORKED EXAMPLE

120 students choose how they travel to school and 54 choose the bus. Find the angle for the bus sector.

1. angle = 54 ÷ 120 × 360°
2. **= 0.45 × 360° = 162°**

WORKED EXAMPLE

In a pie chart a 60° sector shows 18 students. How many students are there in total, and how many are in a 100° sector?

1. 60° is 18 students, so 1° is 18 ÷ 60 = 0.3 students
2. total = 360 × 0.3 = **108 students**
3. **100° sector = 100 × 0.3 = 30 students**

COMMON MISTAKE

A bar chart axis that does **not start at zero**, or that has an uneven scale, makes small differences look huge.

COMMON MISTAKE

Using 100 instead of **360** when working out pie chart angles.

Stem-and-leaf diagrams

Ages of 18 people

stem	leaf
1	2 5 8
2	1 3 3 3 6 9
3	0 2 4 4 7 8
4	1 5 6

Key: 2 | 3 means 23 years

Read the diagram

18 values, in order:
12 15 18 21 23 23 23 26 29
30 32 34 34 37 38 41 45 46

Median: (9th + 10th) ÷ 2
= (29 + 30) ÷ 2 = 29.5
Mode = 23 (appears 3 times)
Range = 46 − 12 = 34

Write the leaves in order, one digit each, and always give a key.

The stem holds the tens and the leaves hold the units. The diagram is also a sideways bar chart: the longest row shows where most of the values are.

- **Order the leaves** from smallest to largest, going away from the stem.
- **Write a key** such as "2 | 3 means 23 years". Without it you lose a mark.
- **Every leaf is one digit.** Keep the leaves evenly spaced so the shape is clear.
- **Back-to-back diagrams** share one stem, with leaves on both sides, so you can compare two sets of data. Read the left side from the stem outwards too.
- **Median:** count in from the ends of the ordered leaves. For 18 values, take the mean of the 9th and 10th.

WORKED EXAMPLE

Use the diagram of 18 ages. How many people are older than 30, and how many are aged 25 or under?

1. the values above 30 are 32, 34, 34, 37, 38, 41, 45 and 46, so **8 people**
2. **the values 25 or under are 12, 15, 18, 21, 23, 23 and 23, so 7 people**

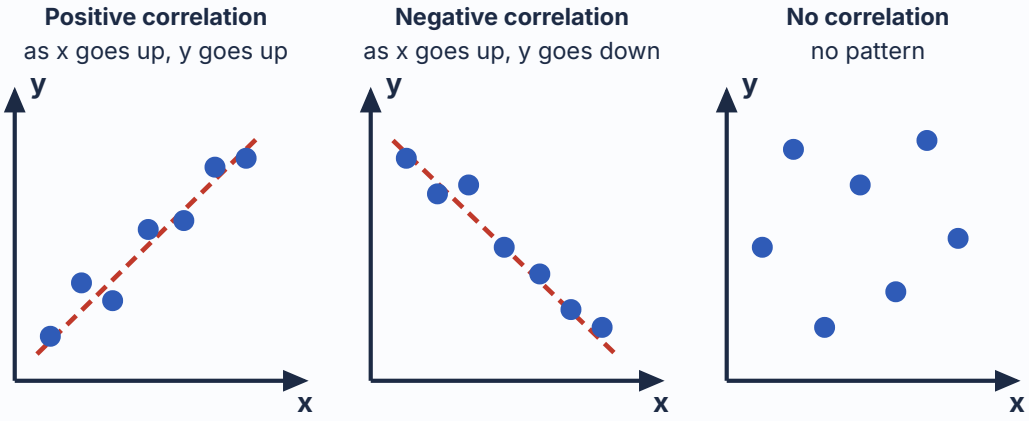
WORKED EXAMPLE

Draw an ordered stem-and-leaf diagram for the marks 34, 28, 45, 31, 28, 39, 42, 27, 36, 31 and find the median and mode.

1. stem 2: 7 8 8
2. stem 3: 1 1 4 6 9
3. stem 4: 2 5
4. key: 3 | 4 means 34
5. ordered data: 27, 28, 28, 31, 31, 34, 36, 39, 42, 45
6. median = (31 + 34) ÷ 2 = **32.5**
7. **modes = 28 and 31 (each appears twice)**

4 Scatter diagrams

Correlation



The closer the points are to a straight line, the stronger the correlation.

Correlation shows a link between the quantities, not that one causes the other.

Correlation describes the link between two sets of data. The closer the points are to a straight line, the stronger it is.

Positive correlation — as one quantity **increases**, the other **increases** too. The points slope **up**.

Negative correlation — as one quantity **increases**, the other **decreases**. The points slope **down**.

No correlation — the points show **no pattern**, so there is no link between the two.

- **Strong or weak:** points close to a line show strong correlation, and points that are more scattered show weak correlation.
- **Describe it in context.** Write "as the temperature rises, ice cream sales rise", not just "positive".
- **Correlation is not causation.** A link does not prove that one thing causes the other. Ice cream sales and sunburn are both caused by hot weather.

WORKED EXAMPLE

Say what type of correlation you expect between (a) height and shoe size (b) age of a car and its value (c) shoe size and exam mark.

1. (a) **positive**, because taller people usually have bigger feet
2. (b) **negative**, because older cars are usually worth less
3. **(c) no correlation, because the two are not linked**

COMMON MISTAKE

Saying "the graph goes up" instead of naming the **type of correlation** and giving the **context**.

Riyaz Sir

Line of best fit



Draw a straight line through the middle of the points, with about the same number above and below it.

Inside the data: reliable (interpolation).
Outside the data: not reliable (extrapolation).

12 hours would give 112, which is impossible out of 100.

Draw one straight line through the middle of the points, with about the same number above and below. Read values from it.

- **Draw it with a ruler.** It does not have to go through the origin, but it should pass through the **mean point** if you have worked it out.
- **Ignore outliers** when you draw the line, and say so if asked.
- **Interpolation** means predicting inside the range of the data, and it is reasonably reliable.
- **Extrapolation** means predicting outside the range of the data. The pattern may not continue, so it is **unreliable**.
- **Gradient meaning:** if the line rises 7 marks for each extra hour, say "about 7 extra marks for each hour of revision". EXT

WORKED EXAMPLE

The diagram shows the line of best fit for hours of revision and test score. Estimate the score for 5.5 hours.

1. draw a vertical line up from 5.5 on the x-axis to the line of best fit
2. go across to the y-axis
3. **score is about 67 (answers from 65 to 69 are accepted)**

WORKED EXAMPLE

The line of best fit passes through (2, 42) and (8, 84). Find its gradient and say what it means.

1. gradient = (84 − 42) ÷ (8 − 2) = 42 ÷ 6 = **7**
2. **it means about 7 extra marks for each extra hour of revision**

WORKED EXAMPLE

Why should the line of best fit not be used to predict the score for 12 hours?

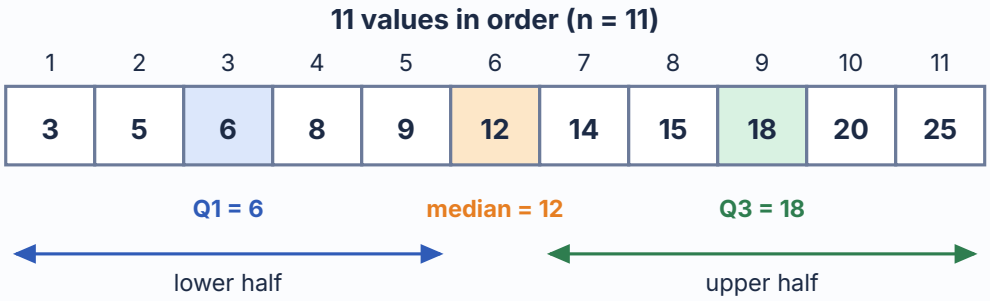
1. 12 hours is **outside the range** of the data (extrapolation)
2. the line would give a score above 100
3. **so the prediction is unreliable**

EXAM TIP

Always say "estimate". A prediction from a line of best fit is an estimate, so a range of answers is accepted.

Quartiles and the interquartile range

EXTENDED



Position of Q1 = (n + 1) ÷ 4 = 12 ÷ 4 = 3rd value
Position of the median = (n + 1) ÷ 2 = 6th value
Position of Q3 = 3(n + 1) ÷ 4 = 9th value

IQR = Q3 − Q1 = 18 − 6 = 12

The quartiles split the ordered data into four equal parts. The IQR is the range of the middle half.

Lower quartile Q1 — the value **a quarter of the way** along the ordered data, with 25% of values below it.

Upper quartile Q3 — the value **three quarters of the way** along, with 25% of values above it.

Interquartile range — **IQR = Q3 − Q1**, the spread of the **middle 50%** of the data.

$$\text{Q1 position} = \frac{n + 1}{4}$$

in the ordered list of n values

$$\text{Q3 position} = \frac{3(n + 1)}{4}$$

in the ordered list of n values

- **If the position is not a whole number,** go that fraction of the way between the two values. For example, the 3.5th value is halfway between the 3rd and 4th.
- **The IQR is better than the range** when there are outliers, because it ignores the extreme values at each end.
- **Order first,** as always. Then count positions from the start of the list.

WORKED EXAMPLE

Find the median, lower quartile, upper quartile and IQR of 4, 7, 9, 10, 12, 13, 16, 18, 21, 23, 25, 30, 31, 35, 40.

1. n = 15 so the data has 15 values already in order
2. median is the (15 + 1) ÷ 2 = 8th value = **18**
3. Q1 is the (15 + 1) ÷ 4 = 4th value = **10**
4. Q3 is the 3 × 16 ÷ 4 = 12th value = **30**
5. **IQR = 30 − 10 = 20**

WORKED EXAMPLE

The data 2, 5, 6, 9, 11, 14, 20 is in order. Find the IQR.

1. n = 7, so Q1 is the 8 ÷ 4 = 2nd value = 5
2. Q3 is the 3 × 8 ÷ 4 = 6th value = 14
3. **IQR = 14 − 5 = 9**

COMMON MISTAKE

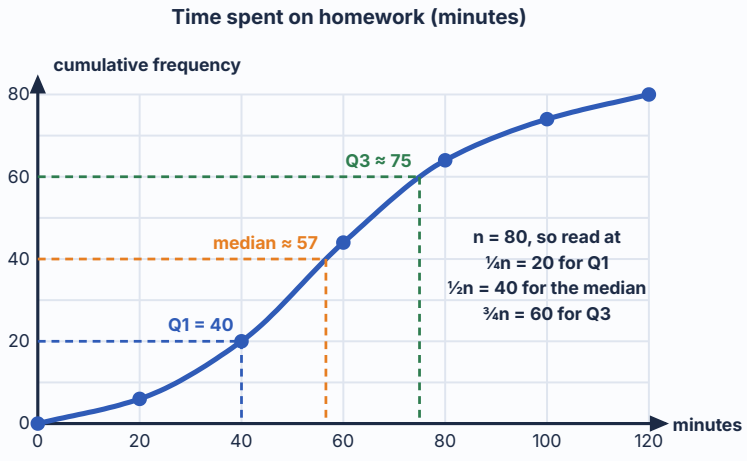
Finding the quartiles of data that is **not in order**.

COMMON MISTAKE

Working out the **position** and then writing it as the answer. The position is only where to look. The answer is the **value** at that position.

Cumulative frequency

EXTENDED



Plot at the upper class boundary, then join with a smooth curve.

IQR ≈ 75 − 40 = 35 minutes

Plot each running total at the upper boundary of its class. Then read across from ¼n, ½n and ¾n to find the quartiles.

Cumulative frequency — the **running total** of the frequencies, which shows how many values are **less than or equal to** a given value.

$$\text{median at } \frac{n}{2}$$

read across from half of the total frequency

$$\text{Q1 at } \frac{n}{4} \text{ and Q3 at } \frac{3n}{4}$$

on the cumulative frequency axis

- **Plot at the upper class boundary,** such as (20, 6) for the class $0 < t \leq 20$, **not** at the middle of the class.
- **Start at the lower boundary of the first class** with a cumulative frequency of 0. Join the points with a **smooth curve**.
- **On the graph use n ÷ 2, not (n + 1) ÷ 2.** The graph reads at ¼n, ½n and ¾n.
- **How many are more than a value?** Read the cumulative frequency at that value, then subtract it from n.
- **Readings from a curve are approximate,** so answers within a small tolerance score the marks. Always show your read-off lines.

WORKED EXAMPLE

The classes for homework time (minutes) are $0 < t \leq 20$: 6, $20 < t \leq 40$: 14, $40 < t \leq 60$: 24, $60 < t \leq 80$: 20, $80 < t \leq 100$: 10, $100 < t \leq 120$: 6. Write the cumulative frequencies.

1. 6
2. 6 + 14 = 20
3. 20 + 24 = 44
4. 44 + 20 = 64
5. 64 + 10 = 74
6. **74 + 6 = 80, which equals the total frequency**

WORKED EXAMPLE

Use the graph (n = 80) to estimate the median, the quartiles and the IQR.

1. Q1: ¼ × 80 = 20, so read across at 20 and down to **40**
2. median: ½ × 80 = 40, so read across at 40 and down to about **57**
3. Q3: ¾ × 80 = 60, so read across at 60 and down to about **75**
4. **IQR = 75 − 40 = 35 minutes**

WORKED EXAMPLE

Use the same graph to estimate how many students spent more than 90 minutes.

1. at 90 minutes the cumulative frequency is about 70
2. **students above 90 minutes = 80 − 70 = about 10**

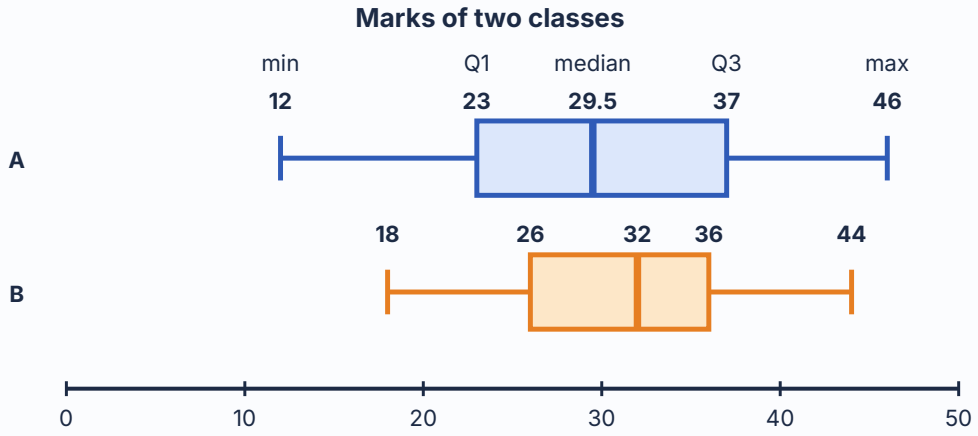
COMMON MISTAKE

Plotting at the **middle** of the class, or plotting the **frequency** instead of the cumulative frequency.

Riyaz Sir

Box plots

EXTENDED



Class B: higher median (32 compared with 29.5) and smaller IQR (10 compared with 14). So B did slightly better and was more consistent.

A box plot shows five numbers: the smallest, Q1, the median, Q3 and the largest. Each section holds a quarter of the data.

Five-number summary — the **minimum, lower quartile, median, upper quartile and maximum**. They are all you need to draw a box plot.

- **The box** goes from Q1 to Q3, so its length is the **IQR**. The line inside the box marks the **median**.
- **The whiskers** reach to the smallest and largest values, so the full length is the **range**.
- **Each section** (whisker, half box, half box, whisker) holds **25%** of the data.
- **Use a scale** and draw all box plots against the same axis if you are comparing them.
- **To compare two sets of data,** write one comment on the **average** (the medians) and one on the **spread** (the IQRs or ranges), both in context.

WORKED EXAMPLE

Compare Class A (min 12, Q1 23, median 29.5, Q3 37, max 46) with Class B (min 18, Q1 26, median 32, Q3 36, max 44).

1. IQR of A = 37 − 23 = 14 and IQR of B = 36 − 26 = 10
2. **average:** Class B has a higher median (32 compared with 29.5), so B scored higher on average
3. **spread: Class B has a smaller IQR (10 compared with 14), so B's marks were more consistent**

WORKED EXAMPLE

A box plot has minimum 5, Q1 12, median 20, Q3 28 and maximum 45. Find the range and IQR, and the number out of 80 students who scored between 12 and 28.

1. range = 45 − 5 = **40**
2. IQR = 28 − 12 = **16**
3. **between Q1 and Q3 is 50% of the students, so 50% of 80 = 40 students**

COMMON MISTAKE

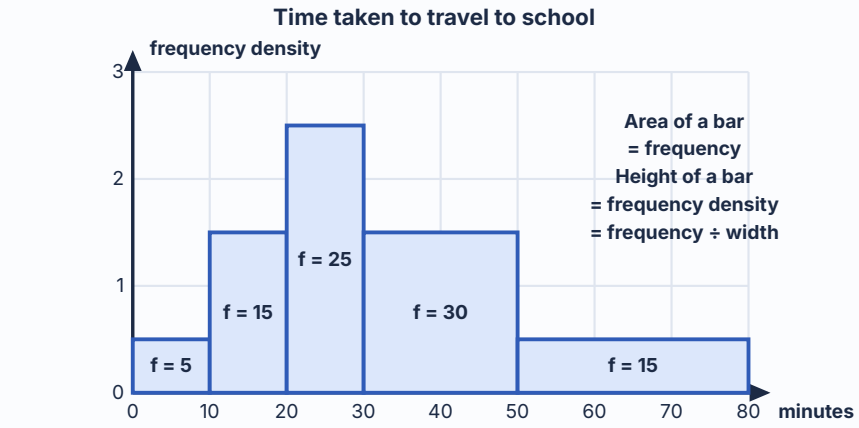
Saying "the box plot with the **longer box** is better". A long box means a **larger IQR**, which means **less consistent**.

EXAM TIP

"Compare" needs two comments using words such as "higher", "smaller" and "more consistent", each with a number and the context.

Histograms

EXTENDED



width	10	10	10	20	30
frequency	5	15	25	30	15
freq. density	**0.5**	**1.5**	**2.5**	**1.5**	**0.5**

The area of each bar is the frequency, so the height is the frequency divided by the class width.

Frequency density — **frequency ÷ class width**. It is the height of the bar in a histogram.

$$\text{frequency density} = \frac{\text{frequency}}{\text{class width}}$$

this is the height of the bar

$$\text{frequency} = \text{frequency density} \times \text{class width}$$

the area of the bar

- **The vertical axis is frequency density,** not frequency. Label it that way.
- **Use a histogram when the class widths are unequal.** If all widths are the same, the height is proportional to the frequency, but you still use the same rule.
- **Bars touch each other,** because the data is continuous. There are no gaps.
- **Estimate within a bar** by taking a fraction of its area, assuming the values are spread evenly across the class.

WORKED EXAMPLE

Use the histogram to find the frequency of the class $30 < t \leq 50$.

1. frequency density = 1.5 and class width = 20
2. **frequency = 1.5 × 20 = 30**

WORKED EXAMPLE

Estimate the number of students with times between 40 and 50 minutes.

1. the bar is 1.5 high
2. the width from 40 to 50 is 10
3. **frequency = 1.5 × 10 = 15**

WORKED EXAMPLE

Estimate the mean time using the frequencies 5, 15, 25, 30 and 15 for the classes 0 to 10, 10 to 20, 20 to 30, 30 to 50 and 50 to 80.

1. midpoints are 5, 15, 25, 40 and 65
2. Σ(f × midpoint) = 25 + 225 + 625 + 1200 + 975 = 3050
3. Σf = 90
4. **estimated mean = 3050 ÷ 90 = 33.9 minutes**

WORKED EXAMPLE

Draw the histogram for $0 < x \leq 10$: 8, $10 < x \leq 30$: 30, $30 < x \leq 40$: 18 and $40 < x \leq 60$: 12.

1. frequency densities: 8 ÷ 10 = 0.8
2. 30 ÷ 20 = 1.5
3. 18 ÷ 10 = 1.8
4. 12 ÷ 20 = **0.6**
5. **draw bars of those heights, with widths 10, 20, 10 and 20 and no gaps**

COMMON MISTAKE

Plotting **frequency** on the vertical axis for unequal classes. The bars will have the wrong shape.

COMMON MISTAKE

Forgetting to **divide by the class width**, so that wide classes look far too tall.

EXAM TIP

A quick check: add up all the bar areas (height × width). The total must equal the total frequency.

Test yourself — cover the answers first

QUICK CHECK

1. Find the mean of 4, 8, 9, 11, 13.
2. Find the median of 7, 2, 9, 4, 12, 3.
3. Find the mode of 5, 7, 5, 8, 7, 5, 9.
4. Find the range of 14, 3, 21, 8.
5. The mean of 6 numbers is 9. What is their total?
6. Which average is best for 12, 13, 13, 14, 80? Give a reason.
7. A frequency table has x = 1, 2, 3 with f = 2, 5, 3. Find the mean.
8. For the same table find the median.
9. What is the midpoint of the class $20 < x \leq 30$?
10. Why is a mean from a grouped table only an estimate?
11. In a pie chart, 45 of 180 students walk to school. What is the angle for walking?
12. What goes wrong when a bar chart's vertical axis does not start at 0?
13. In a stem-and-leaf diagram with key 3 | 4 = 34, what does the row 5 | 0 2 7 show?
14. What type of correlation do you expect between temperature and sales of hot chocolate?
15. What type of correlation do you expect between the age of a car and its value?
16. Why is a prediction from outside the range of the data unreliable?
17. Does correlation between two quantities prove that one causes the other?
18. EXT Find Q1, Q3 and the IQR of 3, 5, 6, 8, 9, 12, 14.
19. EXT A cumulative frequency graph has n = 120. At what cumulative frequency do you read the median?
20. EXT For the same graph, where do you read Q1 and Q3?
21. EXT What is the greatest cumulative frequency in a table?
22. EXT A class $10 < x \leq 30$ has frequency 18. Find the frequency density.
23. EXT A histogram bar has width 5 and height 3. What is the frequency?
24. EXT What percentage of the data lies between the median and Q3?
25. EXT A box plot has minimum 4, Q1 9, median 15, Q3 22 and maximum 40. Find the range and the IQR.

EXAM-STYLE

26. The ages of nine club members are 12, 15, 15, 16, 18, 20, 21, 24 and 30. Find the mean, median, mode and range. [4]
27. The mean of five numbers is 14. A sixth number is added and the mean of the six numbers is 15. Find the sixth number. [3]
28. The number of pets owned by 30 students is 0 pets: 6, 1 pet: 9, 2 pets: 10, 3 pets: 4, 4 pets: 1. Find the mean, the median and the mode. [5]
29. The masses of 40 parcels are $40 < m \leq 50$: 8, $50 < m \leq 60$: 14, $60 < m \leq 70$: 10 and $70 < m \leq 80$: 8. Estimate the mean mass and give the modal class. [4]
30. 72 people were asked how they travel to work: bus 20, car 28, walk 16 and cycle 8. Calculate the angle of each sector of a pie chart. [3]
31. Draw a stem-and-leaf diagram for 23, 31, 18, 27, 35, 22, 29, 31, 40, 26 and find the median. [4]
32. A line of best fit for practice hours (x) and score (y) passes through (2, 15) and (10, 47). Find its gradient, explain it, and estimate the score for 6 hours. [4]
33. EXT Find the median, lower quartile, upper quartile and IQR of 14, 17, 19, 22, 25, 26, 29, 31, 34, 38, 45. [4]

34. **EXT** The times (minutes) of 50 runners are $0 < t \le 10$: 4, $10 < t \le 20$: 11, $20 < t \le 30$: 17, $30 < t \le 40$: 12 and $40 < t \le 50$: 6. Write the cumulative frequencies and say where the median is read from. [3]

35. **EXT** A histogram has bars $0 < x \le 10$ with height 1.2, $10 < x \le 20$ with height 2.6, $20 < x \le 40$ with height 1.8 and $40 < x \le 60$ with height 0.5. Find the total frequency. [4]

36. **EXT** Complete the frequency densities for $0 < a \le 10$: 20, $10 < a \le 30$: 50, $30 < a \le 50$: 30 and $50 < a \le 80$: 45. [4]

37. **EXT** Class X has median 56 and IQR 20. Class Y has median 62 and IQR 12. Compare the two classes. [2]

Riyaz Sir

Show answers

1. 9
2. 5.5 (in order 2, 3, 4, 7, 9, 12)
3. 5
4. 18
5. 54
6. the median (13), because 80 is an outlier that pulls the mean up to 26.4
7. 2.1
8. 2
9. 25
10. the exact values are unknown, so we assume every value is at the midpoint of its class
11. 90°
12. the bars are no longer in proportion, so differences look bigger than they are
13. the values 50, 52 and 57
14. negative
15. negative
16. the pattern may not continue (extrapolation)
17. no, there may be a third factor or just coincidence
18. Q1 = 5, Q3 = 12, IQR = 7
19. 60
20. at 30 and 90
21. the total frequency n
22. 0.9
23. 15
24. 25%
25. range 36, IQR 13
26. total = 171, so mean = 171 ÷ 9 = 19
    median = 5th value = 18
    mode = 15
    range = 30 − 12 = 18
27. total of five = 5 × 14 = 70
    total of six = 6 × 15 = 90
    sixth number = 90 − 70 = 20
28. Σ(f × x) = 0 + 9 + 20 + 12 + 4 = 45, so mean = 45 ÷ 30 = 1.5
    cumulative frequency: 6, 15, 25, 29, 30
    the 15th value is 1 and the 16th value is 2
    median = (1 + 2) ÷ 2 = 1.5
    mode = 2

29. midpoints 45, 55, 65, 75
    Σ(f × midpoint) = 360 + 770 + 650 + 600 = 2380
    mean = 2380 ÷ 40 = 59.5 kg
    modal class = 50 < m ≤ 60

30. bus = 20 ÷ 72 × 360 = 100°
    car = 28 ÷ 72 × 360 = 140°
    walk = 16 ÷ 72 × 360 = 80°
    cycle = 8 ÷ 72 × 360 = 40° (total 360°)

31. stem 1: 8
    stem 2: 2 3 6 7 9
    stem 3: 1 1 5
    stem 4: 0
    key: 2 | 3 means 23
    ordered data: 18, 22, 23, 26, 27, 29, 31, 31, 35, 40
    median = (27 + 29) ÷ 2 = 28

32. gradient = (47 − 15) ÷ (10 − 2) = 32 ÷ 8 = 4
    it means about 4 extra marks for each extra hour of practice
    at 6 hours the score is 15 + 4 × 4 = 31
    an estimate of about 31

33. n = 11
    median = 6th value = 26
    Q1 = 3rd value = 19
    Q3 = 9th value = 34
    IQR = 34 − 19 = 15

34. cumulative frequencies: 4, 15, 32, 44, 50
    points are plotted at the upper boundaries 10, 20, 30, 40, 50
    median is read at ½ × 50 = the cumulative frequency of 25

35. 1.2 × 10 = 12
    2.6 × 10 = 26
    1.8 × 20 = 36
    0.5 × 20 = 10
    total = 12 + 26 + 36 + 10 = 84

36. 20 ÷ 10 = 2
    50 ÷ 20 = 2.5
    30 ÷ 20 = 1.5
    45 ÷ 30 = 1.5

37. Class Y has a higher median (62 compared with 56), so Y scored higher on average
    Class Y has a smaller IQR (12 compared with 20), so Y's marks were more consistent

Stuck on a question?

Send it to me on WhatsApp. I read every message.

Share with a friend

Ask Riyaz Sir

Want regular help? Send this to your parent →