

Representing data

Worked solutions

A-level Maths topic practice

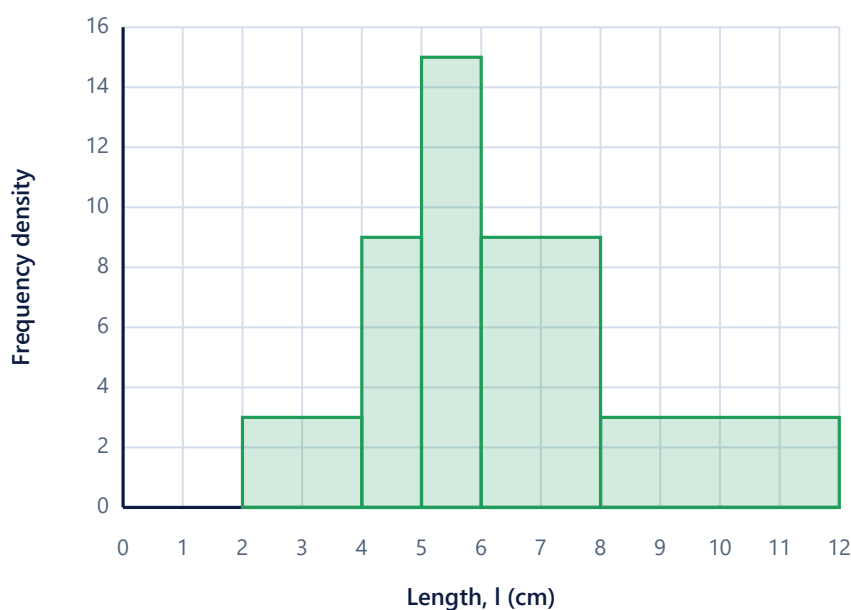


This topic online

- 1 The lengths, l cm, of 60 leaves are summarised in the table.

Length, l (cm)	Frequency
$2 \leq l < 4$	6
$4 \leq l < 5$	9
$5 \leq l < 6$	15
$6 \leq l < 8$	18
$8 \leq l < 12$	12

- a. Calculate the frequency density for each class.
b. Draw a histogram to represent the data on the grid.



(3 marks)

a) Frequency density = $\frac{\text{frequency}}{\text{class width}}$

$$2 \leq l < 4: \frac{6}{2} = 3$$

$$4 \leq l < 5: \frac{9}{1} = 9$$

$$5 \leq l < 6: \frac{15}{1} = 15$$

$$6 \leq l < 8: \frac{18}{2} = 9$$

$$8 \leq l < 12: \frac{12}{4} = 3$$

b) Bar heights are the frequency densities

Bars from 2 to 4, 4 to 5, 5 to 6, 6 to 8 and 8 to 12

Heights 3, 9, 15, 9, 3

No gaps between the bars

Label the vertical axis frequency density

2

The box plot summarises the reaction times, in hundredths of a second, of a group of drivers in a simulator.

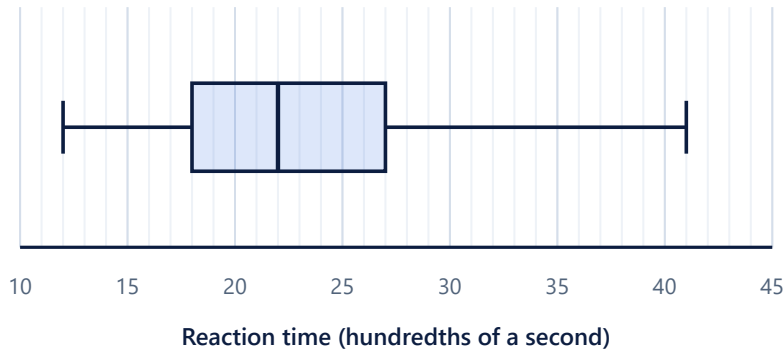
a. Write down the median reaction time.

b. Find the interquartile range.

An outlier is a value greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$.

c. Show that the longest reaction time is an outlier.

d. State one reason why a value in a data set might be an outlier.



(4 marks)

a) The line inside the box

$$\text{Median} = 22$$

b) Read the quartiles from the box

$$Q_1 = 18$$

$$Q_3 = 27$$

$$\text{IQR} = 27 - 18$$

$$\text{IQR} = 9$$

c) Work out the upper limit

$$Q_3 + 1.5 \times \text{IQR} = 27 + 1.5 \times 9$$

$$Q_3 + 1.5 \times \text{IQR} = 27 + 13.5$$

$$Q_3 + 1.5 \times \text{IQR} = 40.5$$

$$\text{Longest time} = 41$$

$$41 > 40.5$$

$$\text{So } 41 \text{ is an outlier}$$

d) Any sensible reason

It may have been recorded wrongly,
or the driver may have been distracted.

3

The table shows the number of hours, h , that 70 students spent watching television in one week.

Hours, h	Frequency
$0 \leq h < 5$	10
$5 \leq h < 10$	18
$10 \leq h < 15$	22
$15 \leq h < 25$	14
$25 \leq h < 40$	6

A histogram is drawn to represent these data. The bar for $10 \leq h < 15$ is 2 cm wide and 8.8 cm tall.

a. Find the width and the height of the bar for $15 \leq h < 25$

b. Estimate the number of students who watched between 12 and 20 hours of television.

(5 marks)

a) Width: 5 hours is 2 cm

$$1 \text{ hour} = \frac{2}{5} = 0.4 \text{ cm}$$

$$\text{Width} = 10 \times 0.4 = 4 \text{ cm}$$

Area is proportional to frequency

$$\text{Area for 22 students} = 2 \times 8.8 = 17.6 \text{ cm}^2$$

$$1 \text{ student} = \frac{17.6}{22} = 0.8 \text{ cm}^2$$

$$\text{Area for 14 students} = 14 \times 0.8 = 11.2 \text{ cm}^2$$

$$\text{Height} = \frac{11.2}{4}$$

Width 4 cm, height 2.8 cm

b) Take the right fraction of each class

$$12 \text{ to } 15: \frac{3}{5} \times 22 = 13.2$$

$$15 \text{ to } 20: \frac{5}{10} \times 14 = 7$$

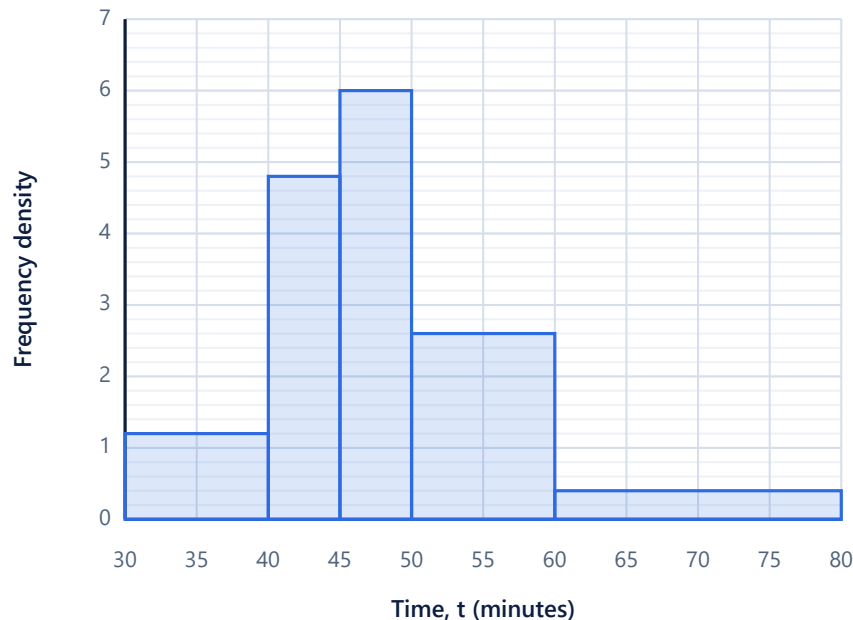
$$\text{Total} = 13.2 + 7 = 20.2$$

About 20 students

4

The histogram shows the times, t minutes, taken by the runners in a 10 km race.

- Find the number of runners in the race.
- Estimate the number of runners who took less than 48 minutes.
- Use linear interpolation to estimate the median time.
- Estimate the mean time.



(7 marks)

a) Frequency = frequency density $\times$ class width

$$30 \text{ to } 40: 1.2 \times 10 = 12$$

$$40 \text{ to } 45: 4.8 \times 5 = 24$$

$$45 \text{ to } 50: 6 \times 5 = 30$$

$$50 \text{ to } 60: 2.6 \times 10 = 26$$

$$60 \text{ to } 80: 0.4 \times 20 = 8$$

$$\text{Total} = 12 + 24 + 30 + 26 + 8$$

100 runners

b) 45 to 48 is 3 of the 5 minutes in its class

$$12 + 24 + \frac{3}{5} \times 30 = 54$$

About 54 runners

c) Median is the 50th time

Cumulative frequency: 12, 36, 66, ...

50th time is in $45 \leq t < 50$

$$\text{Median} = 45 + \frac{50 - 36}{30} \times 5$$

$$\text{Median} = 47.333\dots$$

$$\text{Median} = 47.3 \text{ minutes (3 s.f.)}$$

d) Use the midpoints

$$t = 35, 42.5, 47.5, 55, 70$$

$$\sum ft = 420 + 1020 + 1425 + 1430 + 560$$

$$\sum ft = 4855$$

$$\text{Mean} = \frac{4855}{100}$$

$$\text{Mean} = 48.55$$

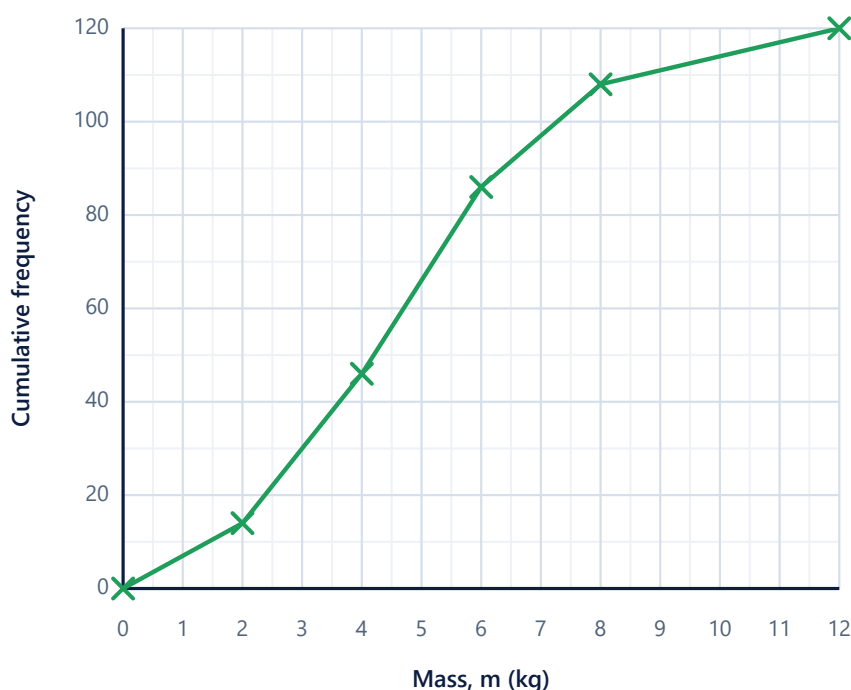
$$\text{Mean} = 48.6 \text{ minutes (3 s.f.)}$$

5

The masses, m kg, of 120 parcels sent from a post office are summarised below.

Mass, m (kg)	Frequency
$0 < m \leq 2$	14
$2 < m \leq 4$	32
$4 < m \leq 6$	40
$6 < m \leq 8$	22
$8 < m \leq 12$	12

- On the grid, draw a cumulative frequency graph for these data, joining the points with straight lines.
- Use your graph to estimate the median and the interquartile range.
- Parcels heavier than 7 kg cost extra to send. Use your graph to estimate the number of parcels that cost extra.



(7 marks)

a) Running totals, plotted at the upper class boundaries

$(0, 0), (2, 14), (4, 46), (6, 86), (8, 108), (12, 120)$

Join with straight lines

b) Median: read across from $\frac{120}{2} = 60$

$$\text{Median} = 4 + \frac{60 - 46}{40} \times 2$$

$$\boxed{\text{Median} \approx 4.7 \text{ kg}}$$

Q1: read across from $\frac{120}{4} = 30$

$$Q1 = 2 + \frac{30 - 14}{32} \times 2 = 3.0$$

Q3: read across from $\frac{3 \times 120}{4} = 90$

$$Q3 = 6 + \frac{90 - 86}{22} \times 2 = 6.36...$$

$$IQR = 6.36 - 3.0$$

$$IQR \approx 3.4 \text{ kg}$$

c) Read up from 7 kg

$$\text{Cumulative frequency at 7 kg} = 86 + \frac{1}{2} \times 22 = 97$$

$$120 - 97 = 23$$

About 23 parcels

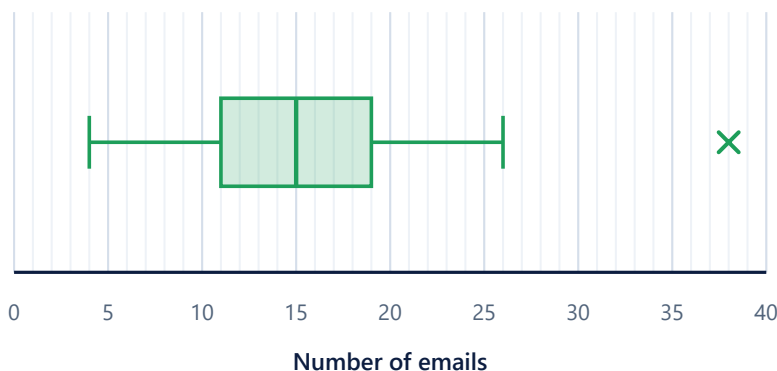
6

Priya records the number of emails she receives each working day for 30 days. Her results are summarised below.

Smallest	Lower quartile	Median	Upper quartile	Largest
4	11	15	19	38

The second largest value is 26. An outlier is a value more than $1.5 \times (Q_3 - Q_1)$ above Q_3 or below Q_1 .

- Show that 38 is the only outlier.
- On the grid, draw a box plot to represent these data.



(4 marks)

a) Outlier Limits

$$IQR = 19 - 11 = 8$$

$$1.5 \times IQR = 12$$

$$\text{Lower Limit} = 11 - 12 = -1$$

$$\text{Upper Limit} = 19 + 12 = 31$$

Smallest value $4 > -1$, so no low outliers

$38 > 31$, so 38 is an outlier

Next largest is $26 < 31$, so it is not

38 is the only outlier

b) Draw the box plot

Box from 11 to 19, line at the median 15

Lower whisker to 4

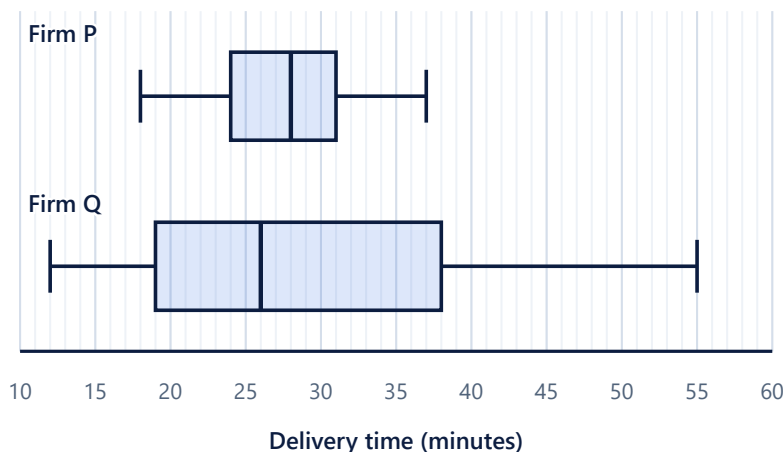
Upper whisker to 26, the largest value that is not an outlier

Mark 38 with a cross

7

A café compares the delivery times, in minutes, of two delivery firms, P and Q . The box plots summarise the times for 40 deliveries by each firm.

- Find the interquartile range for each firm.
- Compare the delivery times of the two firms.
- The café wants a firm whose deliveries are reliably on time. State, giving a reason, which firm it should choose.



(5 marks)

a) $IQR = Q3 - Q1$

P: $IQR = 31 - 24 = 7$

Q: $IQR = 38 - 19 = 19$

P 7 minutes, Q 19 minutes

b) Compare a location and a spread, in context

Median: Q 26 < P 28

so firm Q is slightly quicker on average

IQR: Q 19 > P 7

so firm Q's times are much more variable

c) Reliable means consistent

Firm P: its times vary much less

The times, x seconds, taken by 25 students to solve a puzzle are summarised by

$$\Sigma x = 1150 \quad \Sigma x^2 = 55\,250$$

- a. Find the mean and the standard deviation of the times.

An outlier is a value more than 2 standard deviations from the mean. The fastest time was 24 seconds and the slowest time was 63 seconds.

- b. Show that 24 is an outlier and 63 is not.

- c. The outlier is removed. Find the mean and the standard deviation of the remaining 24 times.

(6 marks)

a) Mean and standard deviation

$$\text{Mean} = \frac{1150}{25} = 46$$

$$\sigma^2 = \frac{55\,250}{25} - 46^2$$

$$\sigma^2 = 2210 - 2116 = 94$$

$$\sigma = \sqrt{94} = 9.695\dots$$

$$\boxed{\text{Mean } 46 \text{ s, } \sigma = 9.70 \text{ s}}$$

b) Limits are mean $\pm 2\sigma$

$$46 - 2 \times 9.695\dots = 26.6$$

$$46 + 2 \times 9.695\dots = 65.4$$

$24 < 26.6$, so 24 is an outlier

$26.6 < 63 < 65.4$, so 63 is not

c) Take 24 out of both sums

$$\Sigma x = 1150 - 24 = 1126$$

$$\Sigma x^2 = 55\,250 - 24^2 = 54\,674$$

$$\text{Mean} = \frac{1126}{24} = 46.916\dots$$

$$\boxed{\text{Mean} = 46.9 \text{ s (3 s.f.)}}$$

$$\sigma^2 = \frac{54\,674}{24} - 46.916\dots^2$$

$$\sigma^2 = 2278.08\dots - 2201.17\dots$$

$$\sigma^2 = 76.909\dots$$

$$\sigma = 8.769\dots$$

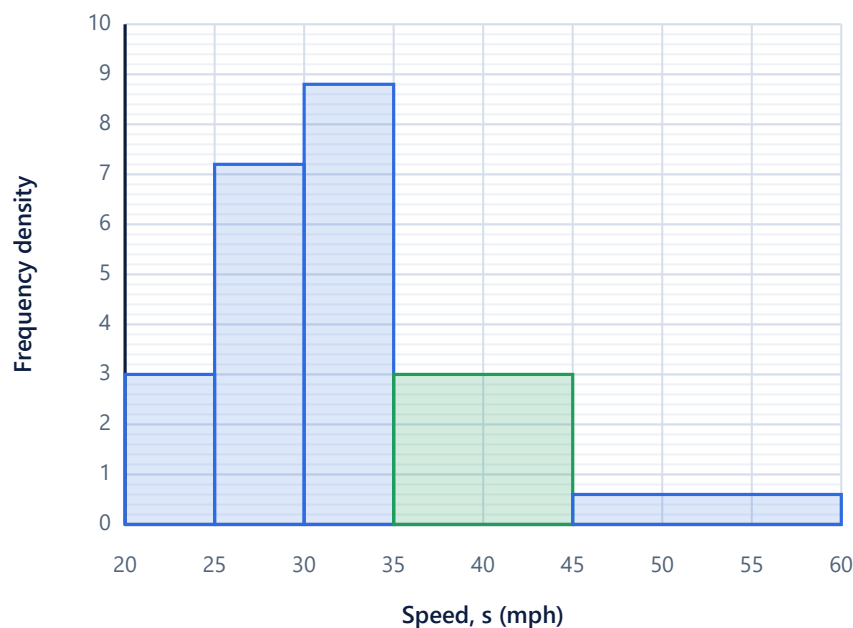
$$\boxed{\sigma = 8.77 \text{ s (3 s.f.)}}$$

9

The speeds, s mph, of cars passing a speed camera were recorded. The table and the histogram are both incomplete.

Speed, s (mph)	Frequency
$20 \leq s < 25$	15
$25 \leq s < 30$	
$30 \leq s < 35$	44
$35 \leq s < 45$	30
$45 \leq s < 60$	9

- Use the histogram to complete the table.
- Complete the histogram.
- Estimate the proportion of these cars travelling at more than 40 mph.



(6 marks)

a) Check the scale with a known class

$$20 \text{ to } 25: \frac{15}{5} = 3$$

matches the bar, so frequency = density $\times$ width

$$25 \text{ to } 30: 7.2 \times 5 = 36$$

36

b) Missing bar for $35 \leq s < 45$

$$\text{Density} = \frac{30}{10} = 3$$

Draw a bar from 35 to 45 of height 3

c) More than 40 mph

$$40 \text{ to } 45: \frac{5}{10} \times 30 = 15$$

$$\text{Number} = 15 + 9 = 24$$

$$\text{Total} = 15 + 36 + 44 + 30 + 9 = 134$$

$$\text{Proportion} = \frac{24}{134}$$

$$\boxed{0.179 \text{ (3 s.f.)}}$$

10

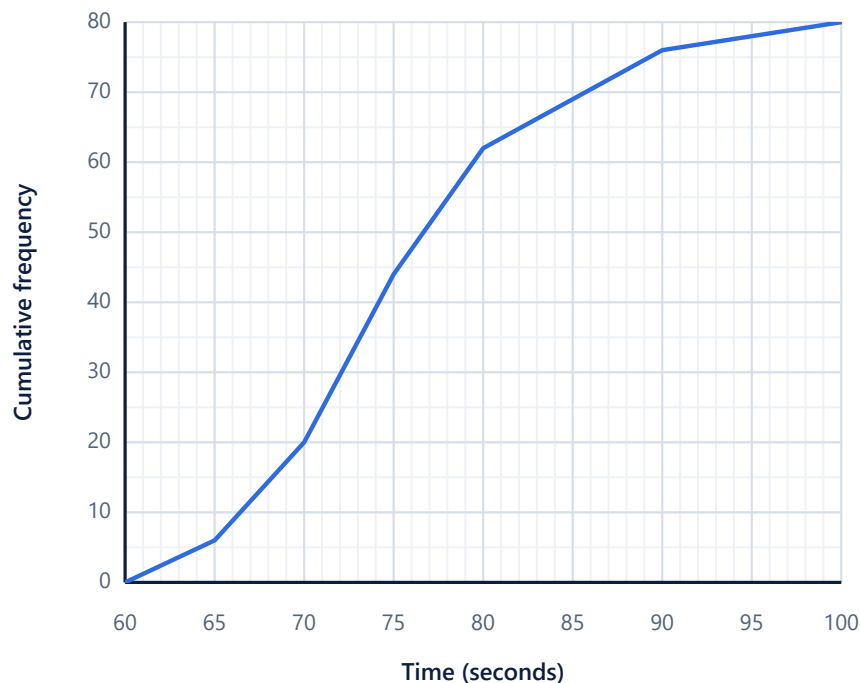
The cumulative frequency graph shows the times, in seconds, taken by 80 Year 12 boys to run 400 m.

a. Use the graph to estimate the median and the interquartile range of the times.

For the Year 12 girls the median time was 81 seconds and the interquartile range was 12 seconds.

b. Compare the times of the boys and the girls.

c. The coach chooses the fastest 25% of the boys for a relay squad. Estimate the slowest time that would get a boy into the squad.



(6 marks)

a) Median: read across from $\frac{80}{2} = 40$

$$\text{Median} = 70 + \frac{40 - 20}{24} \times 5$$

$$\text{Median} = 74.16...$$

$$\boxed{\text{Median} \approx 74 \text{ s}}$$

Q1 at 20 and Q3 at 60

$$Q1 = 70$$

$$Q3 = 75 + \frac{60 - 44}{18} \times 5$$

$$Q3 = 79.44...$$

$$\text{IQR} = 79.44... - 70$$

$$\boxed{\text{IQR} \approx 9.5 \text{ s}}$$

b) Compare a location and a spread, in context

Median: boys 74 < girls 81

so the boys were faster on average

IQR: boys $9.5 <$ girls 12

so the boys' times were more consistent

c) Fastest 25% are below the lower quartile

About 70 s

11

The box plot summarises the times, x minutes, taken by 31 people to assemble a flat-pack bookcase.

For these data $\Sigma x = 589$ and $\Sigma x^2 = 12\,469$

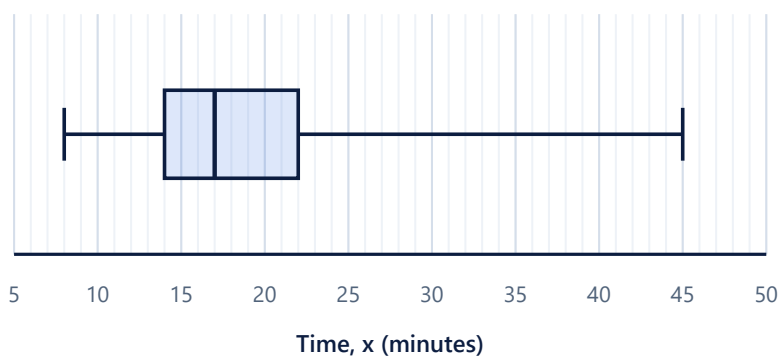
- Find the interquartile range.
- Calculate the mean and the standard deviation of the times.

Ellie defines an outlier as a value more than 2 standard deviations above the mean.

- State, giving a reason, whether there are any outliers in these data using Ellie's definition.

Two more people each take 17 minutes to assemble the bookcase.

- State, giving a reason, the effect of including their times on (i) the median, (ii) the mean.



(8 marks)

a) Read the quartiles

$$IQR = 22 - 14$$

$$IQR = 8 \text{ minutes}$$

b) Mean and standard deviation

$$\text{Mean} = \frac{589}{31} = 19$$

$$\sigma^2 = \frac{12\,469}{31} - 19^2$$

$$\sigma^2 = 402.22... - 361$$

$$\sigma^2 = 41.225...$$

$$\sigma = 6.420...$$

$$\text{Mean } 19, \sigma = 6.42 \text{ minutes}$$

c) Upper limit

$$19 + 2 \times 6.420... = 31.84...$$

$$\text{Longest time} = 45 > 31.8$$

$$\text{Yes, at least one outlier (45)}$$

d)(i) Both new values equal the median

one goes either side of the middle, so the median stays 17

Median unchanged

(ii) Both new values are below the mean of 19

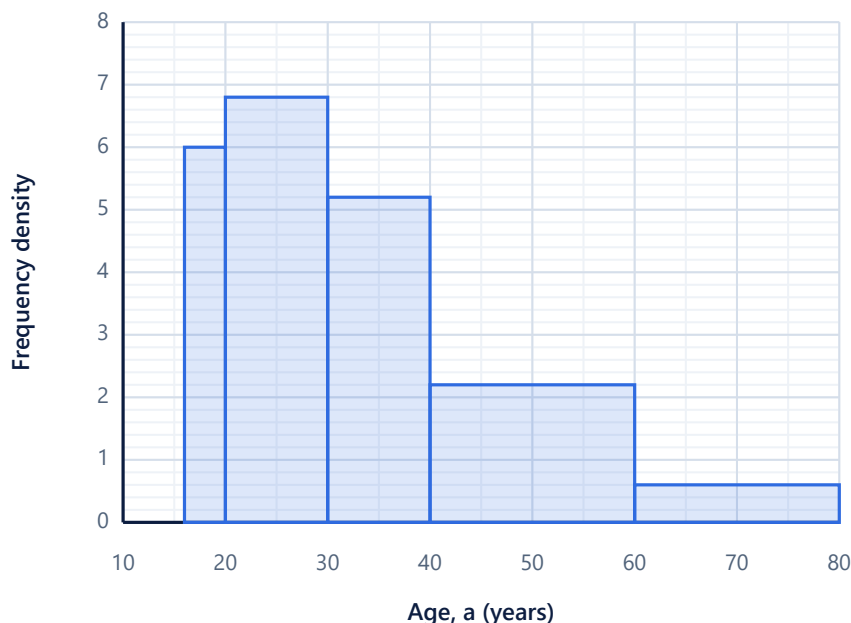
$$\text{New mean} = \frac{589 + 34}{33} = 18.87...$$

Mean decreases

12

The histogram shows the ages, a years, of the 200 members of a gym.

- Estimate the number of members aged between 25 and 45
- Use linear interpolation to estimate the median age.
- Two members are chosen at random. Estimate the probability that both are under 20
- An estimate of the mean age is 35.0 years. Explain, with reference to the histogram, why the mean is greater than the median.



(8 marks)

Frequency = frequency density $\times$ class width

$$16 \text{ to } 20: 6 \times 4 = 24$$

$$20 \text{ to } 30: 6.8 \times 10 = 68$$

$$30 \text{ to } 40: 5.2 \times 10 = 52$$

$$40 \text{ to } 60: 2.2 \times 20 = 44$$

$$60 \text{ to } 80: 0.6 \times 20 = 12$$

Check: total = 200

a) 25 to 30, all of 30 to 40, and 40 to 45

$$= \frac{5}{10} \times 68 + 52 + \frac{5}{20} \times 44$$

$$34 + 52 + 11 = 97$$

About 97 members

b) Median is the 100th age

Cumulative frequency: 24, 92, 144, ...

$$\text{Median} = 30 + \frac{100 - 92}{52} \times 10$$

$$\text{Median} = 31.53\dots$$

$$\text{Median} = 31.5 \text{ years (3 s.f.)}$$

c) Without replacement

$$p = \frac{24}{200} \times \frac{23}{199}$$

$$p = \frac{552}{39800}$$

$$p = 0.0139 \text{ (3 s.f.)}$$

d) Look at the shape

Most members are 16 to 40,

but a tail of older members stretches to 80

These large ages pull the mean up,

but hardly move the median