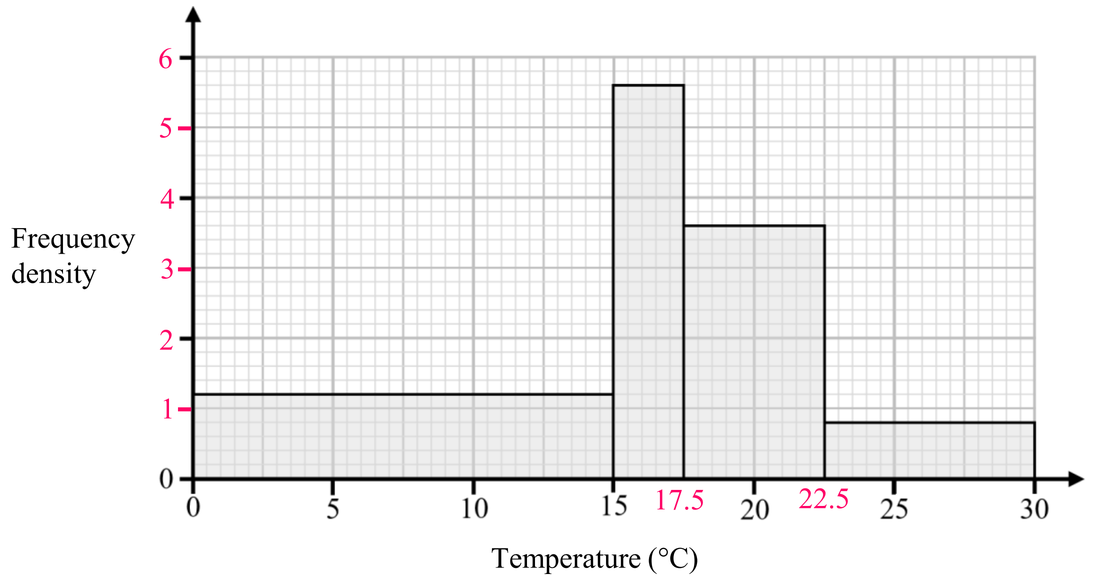
Histograms

REVISE THIS TOPIC

SCAN ME

1 The histogram below gives information about the temperatures, in °C, of some cities on the same day.



18 cities had a temperature less than 15 °C.

(a) Work out the total number of cities represented in the histogram. **(3)**

(b) Use linear interpolation to work out an estimate for the interquartile range of the temperatures. **(4)**

(a) $18 \div 15 = 1.2$

(frequency density for first bar)

$2.5 \times 5.6 = 14$

$5 \times 3.6 = 18$

$7.5 \times 0.8 = 6$

$18 + 14 + 18 + 6 = 56$

(b) $\frac{56}{4} = 14^{\text{th}}$ position

0 — 14 — 18

0 — x — 15

$\frac{14 - 0}{18 - 0} = \frac{x}{15 - 0}$ $x = 11.666\ldots$

$Q_1 = 11.666\ldots$°C

$\frac{3 \times 56}{4} = 42^{\text{nd}}$ position

32 — 42 — 50

17.5 — $17.5 + x$ — 22.5

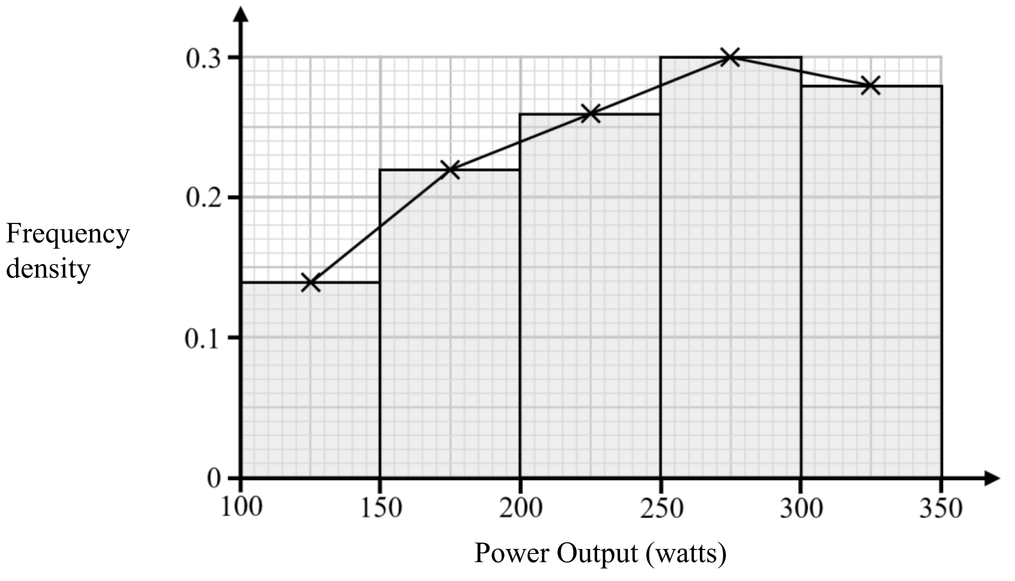
$\frac{42 - 32}{50 - 32} = \frac{x}{22.5 - 17.5}$ $x = 2.777\ldots$

$Q_3 = 20.2777\ldots$°C

$Q_3 - Q_2 = 20.2777\ldots - 11.666\ldots$ **= 8.61 °C**

(Total for Question 1 is 7 marks)

2 The histogram and its frequency polygon below give information about the power outputs, in watts, of 60 cyclists from a cycling club.



(a) Calculate an estimate for the mean power output of the 50 cyclists. **(2)**

(b) Calculate an estimate for the standard deviation of the power outputs of the 50 cyclists. **(2)**

An outlier is any value that falls either

more than 2 × (standard deviation) above the mean or
more than 2 × (standard deviation) below the mean.

The lowest power output of the 50 cyclists was 115 watts.
The highest power output of the 50 cyclists was 339 watts.

(c) Use your answers to parts (a) and (b) to show that, based on the estimates, none of the power outputs are outliers. **(2)**

(d) Use linear interpolation to estimate the median power output.
Give your answer to 1 decimal place. **(4)**

(a) $50 \times 0.14 = 7$

$50 \times 0.22 = 11$

$50 \times 0.26 = 13$

$50 \times 0.3 = 15$

$50 \times 0.28 = 14$

$$\frac{(125 \times 7) + (175 \times 11) + (225 \times 13) + (275 \times 15) + (325 \times 14)}{7 + 11 + 13 + 15 + 14}$$

$= \mathbf{240\ watts}$

(you can type the frequency table using midpoints into your calculator)

(b) $$\sqrt{\frac{(125^2 \times 7) + (175^2 \times 11) + (225^2 \times 13) + (275^2 \times 15) + (325^2 \times 14)}{7 + 11 + 13 + 15 + 14} - 240^2} = \mathbf{66.01\ watts}$$

(c) $240 + 2(66.01) = 372.02$

$240 - 2(66.01) = 107.98$

$115 > 107.98$ $\quad$ $339 < 372.02$ $\quad$ therefore there are no outliers.

(d) $\frac{60}{2} = 30^{th}$ position

18 — 30 — 31

200 — $200 + x$ — 250

$$\frac{30 - 18}{31 - 18} = \frac{x}{250 - 200}$$

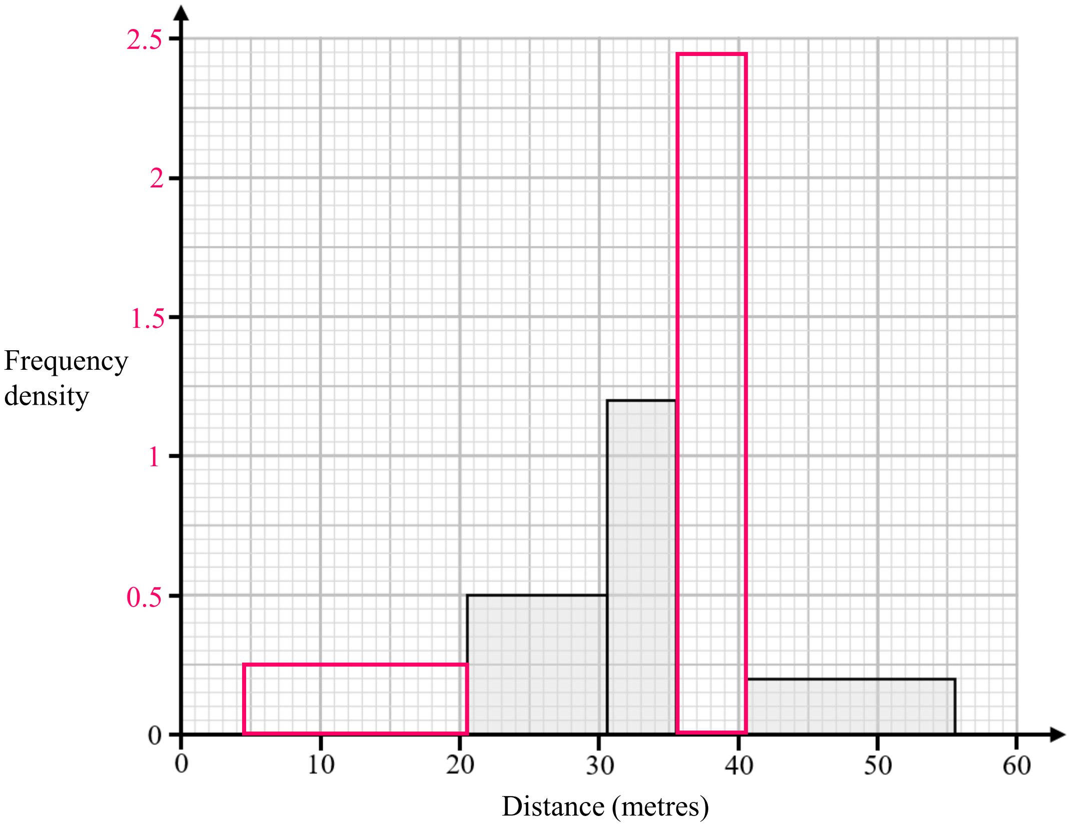
$x = 46.15\ldots$

$\boldsymbol{Q_2} = \mathbf{246.2\ watts}$

(Total for Question 2 is 10 marks)

3 The partially completed table and partially completed histogram below show the distances, to the nearest metre, of javelin throws by athletes in a competition.

Distance (metres)	5 – 20	21 – 30	31 – 35	36 – 40	41 – 55
Frequency	4	5	**7**	12	**6**



(a) Complete the table. **(2)**

(b) Complete the histogram. **(2)**

(a) For 21 – 30 $\quad 20.5 \leq \text{distance} < 30.5$

$5 \div 10 = 0.5$ (frequency density for this bar)

For 31 – 35 $\quad 30.5 \leq \text{distance} < 35.5$

$1.4 \times 5 = 7$ (frequency for this bar)

For 41 – 55 $\quad 40.5 \leq \text{distance} < 55.5$

$0.4 \times 15 = 6$ (frequency for this bar)

(b) For 5 – 20 $\quad 4.5 \leq \text{distance} < 20.5$

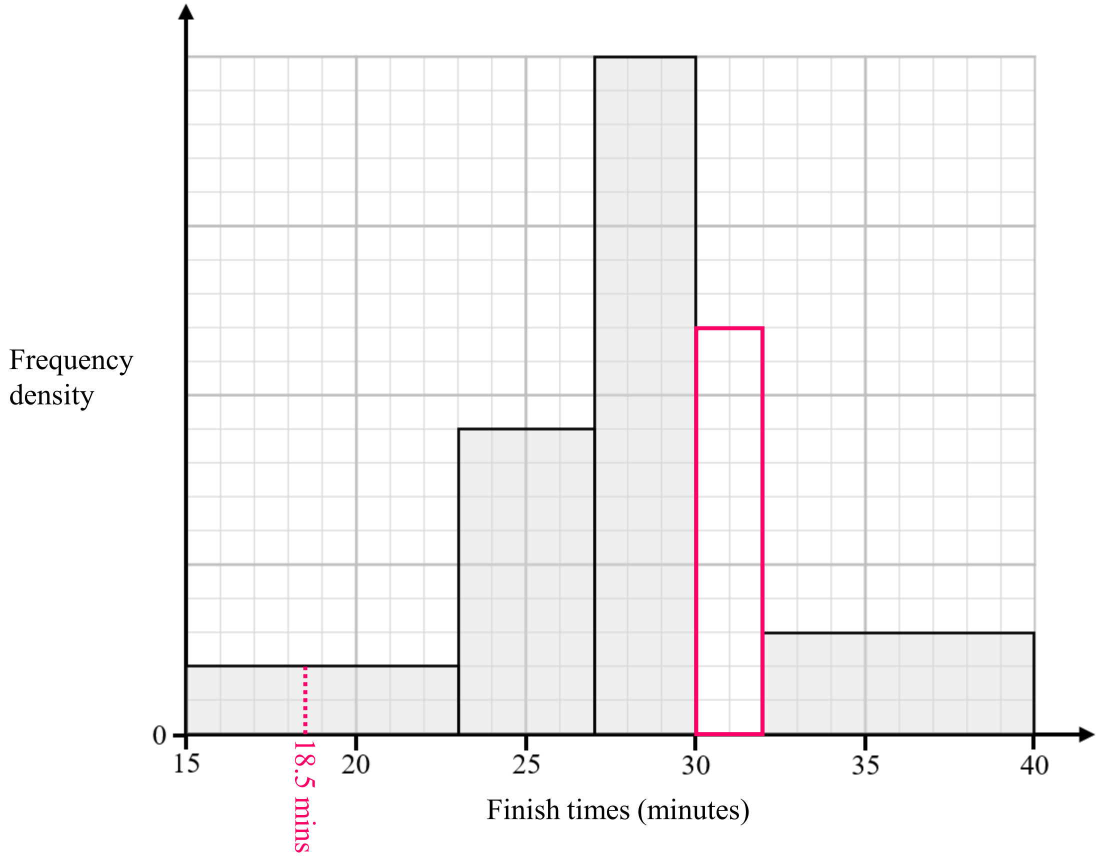
$4 \div 16 = 0.25$ (frequency density for this bar)

For 36 – 40 $\quad 35.5 \leq \text{distance} < 40.5$

$12 \div 5 = 2.4$ (frequency density for this bar)

(Total for Question 3 is 4 marks)

4 The partially completed histogram below gives information about the finish times, in minutes, of runners at a fun run.



All the runners finished with times between 15 and 40 minutes.

10% of the runners finished with a time less than 23 minutes.

(a) Complete the histogram **(3)**

One of the runners is selected at random.

(b) Estimate the probability that this runner had a finish times less than 18 minutes 30 seconds. **(2)**

(c) Use linear interpolation to estimate the interquartile range of the finish times. **(5)**

(a) Area of bar from 15 to 23 = $8 \times 2 = 16$ squares.

16 squares = 10%

160 squares = 100%

Area of other bars: $4 \times 9 = 36$, $3 \times 20 = 60$, $8 \times 3 = 24$

$160 - 16 - 36 - 60 - 24 = 24$ square remaining

$24 \div 2 = 12$ (height of missing bar)

(b) $18.5 - 15 = 3.5$ (width of bar)

Area of bar = $3.5 \times 2 = 7$ (frequency below 18.5 mins)

$\mathbf{\frac{7}{160}}$

(c) $\frac{160}{4} = 40^{th}$ position

16	40	52
23	$23 + x$	27

$\frac{40 - 16}{52 - 16} = \frac{x}{27 - 23}$ $x = 2.666\ldots$

$Q_1 = 25.666\ldots$ mins

$\frac{3 \times 160}{4} = 120^{th}$ position

112	120	136
30	$30 + x$	32

$\frac{120 - 112}{136 - 112} = \frac{x}{32 - 30}$ $x = 0.666\ldots$

$Q_3 = 30.666\ldots$ mins

$Q_3 - Q_2 = 30.666\ldots - 25.666\ldots$ **= 5 minutes**

(Total for Question 4 is 10 marks)

5 The partially completed histogram below gives information about the ages of employees at a large company.

The number of employees aged between 25 and 35 is 80 more than employees aged between 35 and 45.

(a) Work out an estimate for the number of employees aged between 18 and 20. **(3)**

There are a total of 416 employees at the company.
The age of the oldest employee is 75

(b) Complete the histogram by adding a bar for ages between 45 and 75. **(3)**

An outlier is any value that falls either

more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

Given that $Q_1 = 23$ and $Q_3 = 36$

(c) Show that at least one employees age is considered an outlier. **(2)**

The mean age of the employees is 31.7 years, and the standard deviation is 11.8 years.

The CEO of the company assumes that in 1 year's time, the same 416 employees will be working at the company and that there will be no new employees.

(d) Using the CEO's assumption, write down
(i) the mean age of employees at the company in 1 year's time.
(ii) the standard deviation of the ages of the employees in 1 year's time. **(2)**

(a) Area of bar from 25 to 35 = $10 \times 17 = 170$ small squares.

Area of bar from 35 to 45 = $10 \times 7 = 70$ small squares.

$170 - 70 = 100$ small squares difference

100 small squares = 80 employees

1 small square = 0.8 employees

Area of bar from 18 to 25 = $7 \times 25 = 175$ small squares.

$175 \times 0.8 =$ **140 employees**

(b) $170 \times 0.8 = 136$ employees

$70 \times 0.8 = 56$ employees

$416 - 140 - 136 - 56 = 84$ employees remaining

$84 \div 0.8 = 105$ small squares required

Class width = $75 - 45 = 30$

$105 \div 30 = 3.5$ (frequency density)

(c) $36 + 1.5(36 - 23) = 55.5$

$23 - 1.5(36 - 23) = 3.5$

$75 > 55.5$ so it is an outlier.

(d)(i) Mean increases by 1 year to **32.7 years**

(ii) Standard deviation remains the same as the spread has not changed so **11.8 years**

(Total for Question 5 is 10 marks)

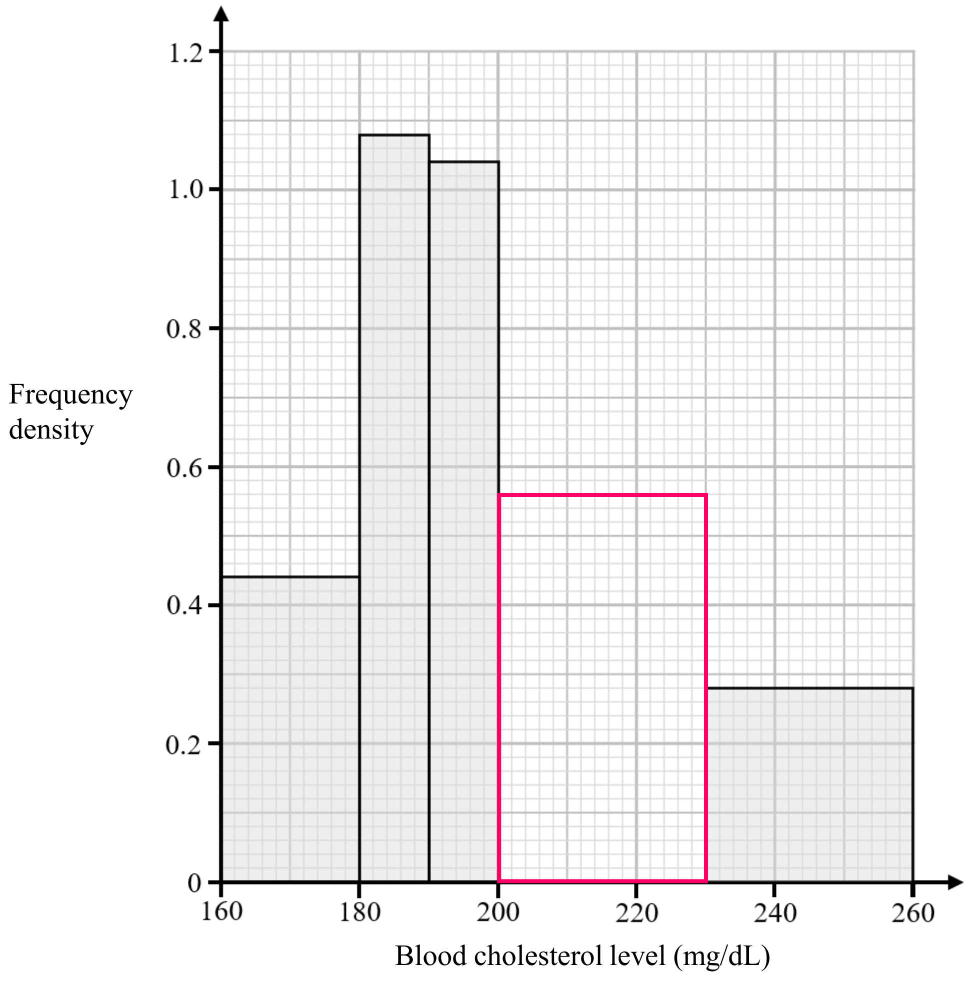
6 A doctor wanted to investigate the blood cholesterol levels of patients at their surgery.
They decide to take a sample of 200 patients.

(a) Identify the population. **(1)**

The doctor asks the next 200 patients who book appointments at the surgery if they will take a blood cholesterol level test.

(b) State the type of sampling technique used by the doctor. **(1)**

The partially completed histogram below shows the blood cholesterol levels of the patients who agreed to do the test.



22 of the patients had a blood cholesterol level between 160 mg/dL and 180 mg/dL.

(c) Work out the number of patients with a blood cholesterol level between 230 mg/dL and 260 mg/dL. **(3)**

60% of the 200 patients who were asked, agreed to take the test.
Of these patients, all recorded a blood cholesterol level between 160 mg/dL and 260 mg/dL.

(d) Complete the histogram by adding the bar for blood cholesterol levels between 200 mg/dL and 230 mg/dL **(3)**

(a) The population is all of the patients who use the doctor's surgery.

(b) Opportunity sampling (or convenience sampling)

(c) For bar between 160 and 200

Classwidth = 200 – 160 = 40

Frequency density = 0.44

$20 \times 0.44 = 8.8$

$22 \div 8.8 = 2.5$

so frequency = $2.5 \times$ area of bar

For bar between 230 and 260

Area = $30 \times 0.28 = 8.4$

Frequency = $2.5 \times 8.4 =$ **21**

(d) $0.6 \times 200 = 120$

$10 \times 1.08 \times 2.5 = 27$ patients

$10 \times 1.04 \times 2.5 = 26$ patients

$160 - 44 - 27 - 26 - 21 = 42$ patients remaining

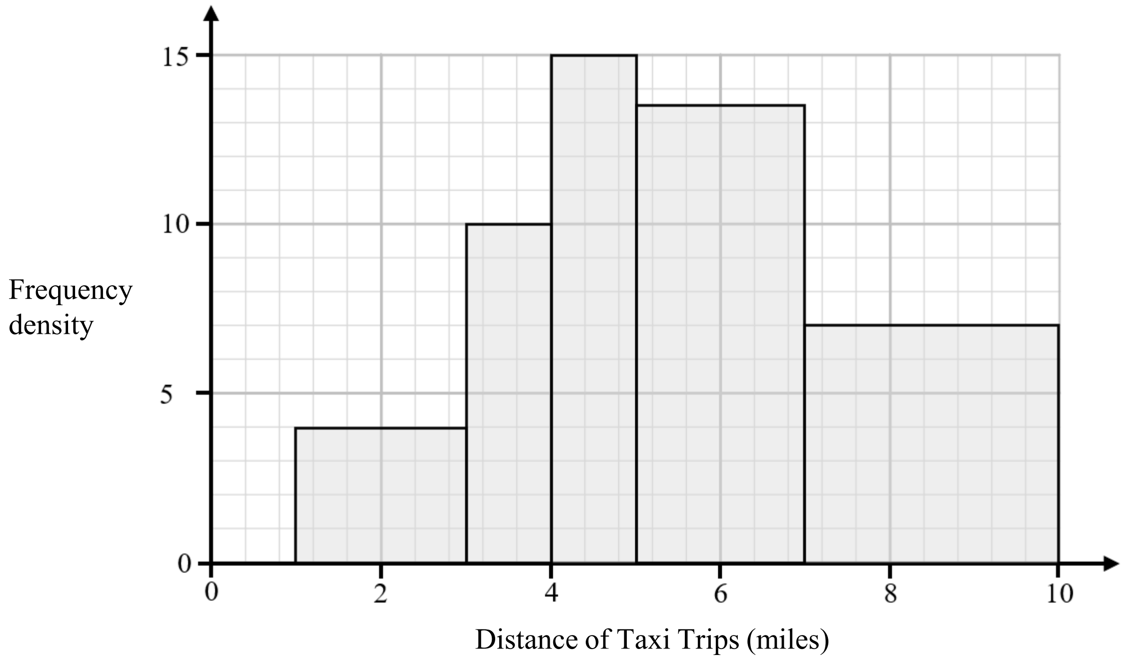
$42 \div 2.5 = 16.8$ (required area of bar)

Class width = 230 – 200 = 30

$16.8 \div 30 = 0.56$ (required frequency density)

(Total for Question 6 is 8 marks)

7 The histogram below shows the distances, in miles, of 81 taxi trips provided by a taxi driver.



The taxi driver only provides trips if the journey distance is between 1 mile and 10 miles.

(a) Calculate the percentage of the taxi trips that were between 3 and 5 miles. **(2)**

The mean distance of the taxi trips is 5.6 miles.
The standard deviation of the distances of the taxi trips is 2.35 miles.

An outlier is any value that falls either

more than 2 × (standard deviation) above the mean or
more than 2 × (standard deviation) below the mean.

(b) Show that none of the distances of the taxi trips were considered outliers. **(2)**

The taxi driver decides to model the **frequency density** for the 81 taxi trips by a curve with equation

$$y = k(1 - x)(x - 10) \qquad 1 \leq x \leq 10$$

where k is a constant.

(c) Find the value of k. **(4)**

(a) Area of each bar:

$2 \times 4 = 8$

$1 \times 10 = 10$

$1 \times 15 = 15$

$2 \times 13.5 = 27$

$3 \times 7 = 21$

$10 + 15 = 25$ (area between 3 and 5 miles)

$8 + 10 + 15 + 27 + 21 = 81$ (total area)

$$\frac{25}{81} \times 100 = 30.86\ \%$$

(b) $5.6 + 2(2.35) = 10.3$

$5.6 - 2(2.35) = 0.9$

$10 < 10.3$ and $1 > 0.9$ so there cannot be any outliers.

(c) $$\int_1^{10} k(1-x)(x-10)\,\mathrm{d}x = 81$$

$$k\int_1^{10} -x^2 + 11x - 10\,\mathrm{d}x = 81$$

$$k\left[-\frac{x^3}{3} + \frac{11x^2}{2} - 10x\right]_1^{10} = 81$$

$$k\left(\left(-\frac{10^3}{3} + \frac{11(10)^2}{2} - 10(10)\right) - \left(-\frac{1^3}{3} + \frac{11(1)^2}{2} - 10(1)\right)\right) = 81$$

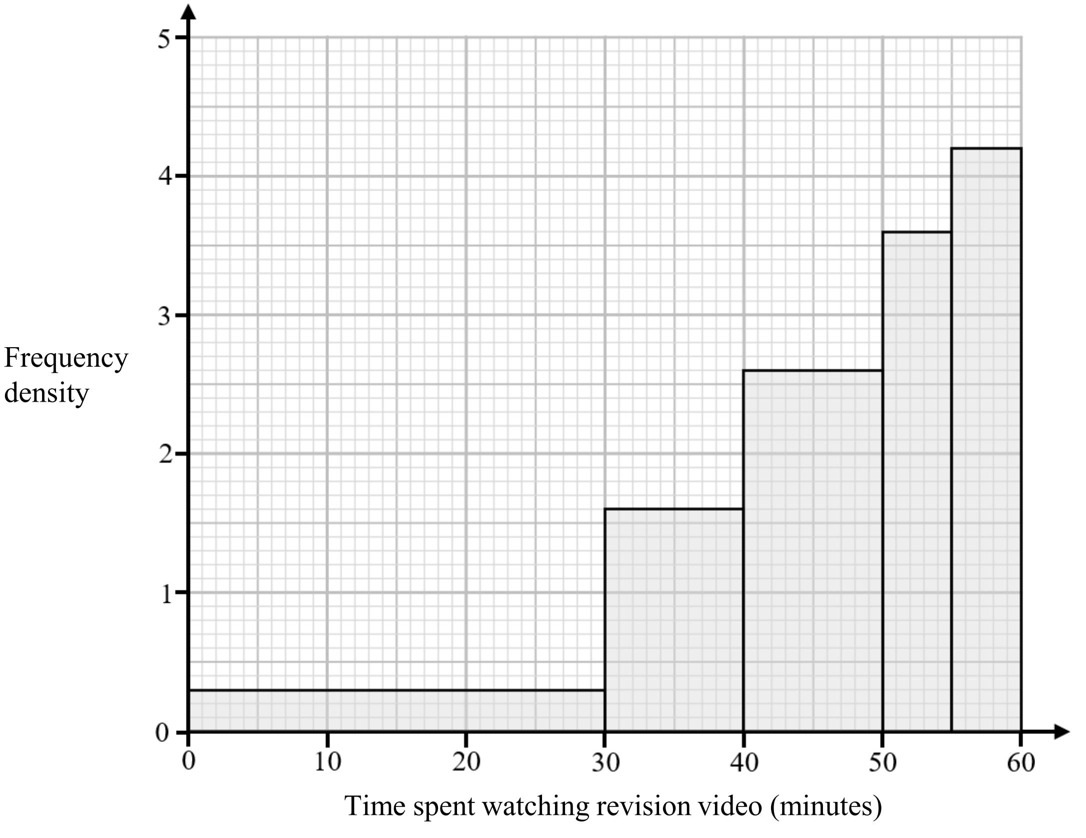
$$\frac{243k}{2} = 81$$

$$k = \frac{81 \times 2}{243} \qquad \boldsymbol{k = \frac{2}{3}}$$

(Total for Question 7 is 8 marks)

8 90 students were asked to watch a 60-minute revision video to prepare for their exams.

The histogram below shows the amount of time each student spent watching the video, up to a maximum of 60 minutes.



(a) Use linear interpolation to work out an estimate for the median time spent watching the video. Give your answer in minutes and seconds, to the nearest second. **(4)**

The class teacher decides to model the **frequency density** for the 90 times by a curve with equation

$$y = \frac{1}{k}x^2 \qquad 0 \leq x \leq 60$$

where k is a constant.

(b) Find the value of k. **(4)**

(a) Area of each bar:

$30 \times 0.3 = 9$

$10 \times 1.6 = 16$

$10 \times 2.6 = 26$

$5 \times 3.6 = 18$

$5 \times 4.2 = 21$

$\frac{90}{2} = 45^{th}$ position

$$\frac{45 - 25}{51 - 25} = \frac{x}{50 - 40}$$

$x = 7.692307\ldots$

$Q_2 = 47.692307\ldots$ mins

$\boldsymbol{Q_2 = 47}$ **mins and 42 seconds**

(b) $\frac{1}{k}\int_0^{60} x^2 \, dx = 90$

$$\frac{1}{k}\left[\frac{x^3}{3}\right]_0^{60} = 90$$

$$\frac{1}{k}\left(\left(\frac{60^3}{3}\right) - \left(\frac{0^3}{3}\right)\right) = 90$$

$$\frac{72000}{k} = 90$$

$$k = \frac{72000}{90}$$

$\boldsymbol{k = 800}$

(Total for Question 8 is 8 marks)