Question 1

Every week an orangutan sanctuary measures the weight of each of its orangutans.

The weights, to the nearest kg, of **ALL** their 18 adult males are listed below:

52, 57, 63, 80, 56, 66, 101, 68, 55, 96, 70, 62, 66, 64, 99, 91, 55, 92

(a) Using a convenience sample of size six, calculate the mean weight of the male orangutans from the data set above.

[2]

(b) Starting from the third data value, take a systematic sample of size six and re-calculate the mean weight of the male orangutans from the data set above.

[2]

(c) Compare your results from parts (a) and (b) and state, with a reason, which sampling method is more reliable.

[2]

a) Just take the first 6 weights:

$$\text{mean} = \frac{52+57+63+80+56+66}{6}$$

$$= \frac{187}{3} = 62.333333\ldots$$

mean = 62.3 kg (1 d.p.)

Note: There are other possible convenience samples! The one used here is just an especially obvious one.

Every week an orangutan sanctuary measures the weight of each of its orangutans.

The weights, to the nearest kg, of **ALL** their 18 adult males are listed below:

52, 57, 63, 80, 56, 66, 101, 68, 55, 96, 70, 62, 66, 64, 99, 91, 55, 92

(a) Using a convenience sample of size six, calculate the mean weight of the male orangutans from the data set above.

[2]

(b) Starting from the third data value, take a systematic sample of size six and re-calculate the mean weight of the male orangutans from the data set above.

[2]

(c) Compare your results from parts (a) and (b) and state, with a reason, which sampling method is more reliable.

[2]

b) $\frac{18}{6} = 3$ take every 3rd weight

$$\text{mean} = \frac{63+66+55+62+99+92}{6}$$

$$= \frac{437}{6} = 72.833333\ldots$$

mean = 72.8 kg (1 d.p.)

Every week an orangutan sanctuary measures the weight of each of its orangutans.

The weights, to the nearest kg, of **ALL** their 18 adult males are listed below:

52, 57, 63, 80, 56, 66, 101, 68, 55, 96, 70, 62, 66, 64, 99, 91, 55, 92

(a) Using a convenience sample of size six, calculate the mean weight of the male orangutans from the data set above.

62.3 kg (1 d.p.)

[2]

(b) Starting from the third data value, take a systematic sample of size six and re-calculate the mean weight of the male orangutans from the data set above.

72.8 kg (1 d.p.)

[2]

(c) Compare your results from parts (a) and (b) and state, with a reason, which sampling method is more reliable.

[2]

c) The sum of the 18 weights is 1293.
So the true mean is:

$\frac{1293}{18} = \frac{431}{6} = 71.8\dot{3}$ kg

The systematic sample mean (72.8) is a lot closer to the true mean here than the convenience sample mean (62.3) is.

Because a systematic sample is generally more random than a convenience sample, it will often give more representative results.

This isn't guaranteed, though! Starting on the first data value in (b) would give a mean of 81, which is not much more accurate than the value in (a). Starting on the second value in (b) would give a mean of 61.6, which is less accurate than the value in (a).

Question 2

A supermarket wants to gather data from its shoppers on how far they have travelled to shop there. One lunchtime an employee is stationed at the door of the shop for half an hour and instructed to ask every customer how far they have travelled.

(a) (i) State the sampling method the employee is using.

(ii) Give one advantage and one disadvantage of using this method.

[3]

(b) State and briefly describe an alternative method of non-random sampling that the employee could use to obtain the required data for a sample of 30 customers.

[2]

a) (i) Opportunity sampling

(ii) This survey will be quick, easy, and inexpensive to carry out.

But it is unlikely to be representative of the shop's customers as a whole – it will only be surveying people who shop during lunchtime on a particular day of the week.

A supermarket wants to gather data from its shoppers on how far they have travelled to shop there. One lunchtime an employee is stationed at the door of the shop for half an hour and instructed to ask every customer how far they have travelled.

(a) (i) State the sampling method the employee is using.

(ii) Give one advantage and one disadvantage of using this method.

[3]

(b) State and briefly describe an alternative method of non-random sampling that the employee could use to obtain the required data for a sample of 30 customers.

[2]

b) Quota sampling. For example, divide the shop's opening hours over a week into a number of 'blocks' and then survey a fixed number of people during each block to create the sample.

Question 3

A pharmacy sells face masks in a variety of sizes. Their sales over a week are recorded in the table below:

	Kids		Adults			
Size	Small	Large	S	M	L	XL
Frequency *f*	29	4	8	24	15	4

(i) Write down the mode for this data.

(ii) Explain why, in this case, the mode from part (i) would not be particularly helpful to the shop owner when reordering masks.

(iii) Given that the shop is open every day of the week, calculate the mean number of masks sold per day.

[4]

(i) Kids small — size with highest frequency

(ii) The mode is a kids size, but the majority of masks sold (51 out of 84) are adult sizes.

(iii) Number sold during week:

$29 + 4 + 8 + 24 + 15 + 4 = 84$

$84 \div 7 =$ 12 masks per day

Question 4

The lengths (l cm) of nine otters, measured to the nearest centimetre by a wildlife research team, are:

76 77 91 65 63 83 92 61 88

Calculate the standard deviation of the nine recorded lengths.

[3]

$$\text{standard deviation} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$\Sigma x = 696$ $\Sigma x^2 = 54998$ $n = 9$

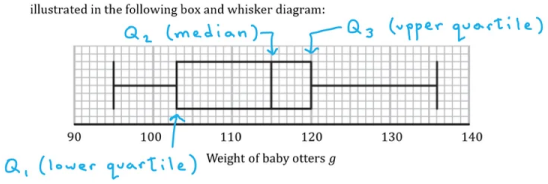
$$\sigma = \sqrt{\frac{54998}{9} - \left(\frac{696}{9}\right)^2} = 11.421227...$$

$\bar{x} = \frac{\Sigma x}{n}$ (mean)

$\sigma = 11.4$ cm (3 s.f.)

Question 5

Jeanette works for a conservation charity who rescue orphaned otters. Over many years she records the weight (g) of each otter when it first arrives. The data is illustrated in the following box and whisker diagram:



(a) Using the box plot above:

(i) Write down the median weight of the otters.

(ii) Write down the lower quartile.

(iii) Find the interquartile range.

[4]

a) (i) median = 115 g

(ii) $Q_1 = 103$ g

(iii) $Q_3 = 120$ g

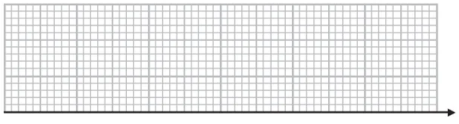
$Q_3 - Q_1 = 120 - 103 = 17$

IQR = 17 g

Otters are then weighed weekly to track their growth. Summary data on the weights (g) of otters after one month is shown in the table below:

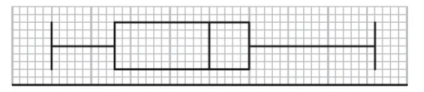
	Weight g
Smallest weight	125
Range	48
Median	152
Upper Quartile	164
Interquartile Range	33

(b) On the grid, draw a box plot for the information given above.



[3]

Jeanette works for a conservation charity who rescue orphaned otters. Over many years she records the weight (g) of each otter when it first arrives. The data is illustrated in the following box and whisker diagram:



Weight of baby otters g

(a) Using the box plot above:

(i) Write down the median weight of the otters.

(ii) Write down the lower quartile.

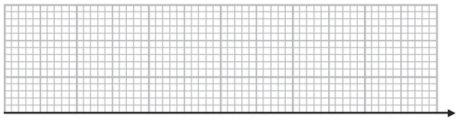
(iii) Find the interquartile range.

[4]

Otters are then weighed weekly to track their growth. Summary data on the weights (g) of otters after one month is shown in the table below:

	Weight g
Smallest weight	125
Range	48
Median	152
Upper Quartile	164
Interquartile Range	33

(b) On the grid, draw a box plot for the information given above.



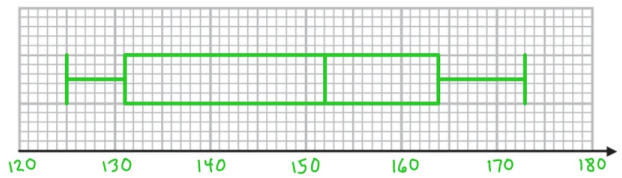
[3]

smallest ↓ range ↓

b) largest weight = 125 + 48 = 173 g

LQ = 164 - 33 = 131 g

↑ UQ ↑ IQR



Question 6

The heights, in metres, of a flock of 20 flamingos are recorded and shown below:

0.4	0.9	1.0	1.0	1.2	1.2	1.2	1.2	1.2	1.2
1.3	1.3	1.3	1.4	1.4	1.4	1.4	1.5	1.5	1.6

An outlier is an observation that falls either more than 1.5 × (interquartile range) above the upper quartile or less than 1.5 × (interquartile range) below the lower quartile.

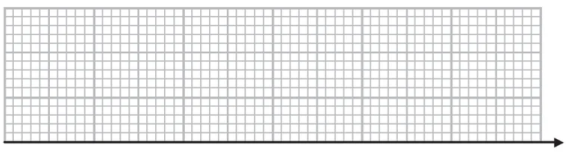
(a) (i) Find the values of Q_1, Q_2 and Q_3.

(ii) Find the interquartile range.

(iii) Identify any outliers.

[4]

(b) Using your answers to part (a), draw a box plot for the data.



[3]

a) (i) Median is $\frac{20+1}{2} = 10.5^{th}$ value:

$Q_2 = \frac{1.2+1.3}{2} = 1.25$ m

$\frac{20}{4} = 5 \Rightarrow Q_1$ is 5.5^{th} value

$20 \times \frac{3}{4} = 15 \Rightarrow Q_3$ is 15.5^{th} value

$Q_1 = \frac{1.2+1.2}{2} = 1.2$ m $\quad Q_3 = \frac{1.4+1.4}{2} = 1.4$ m

(ii) IQR = 1.4 - 1.2 = 0.2 m $\quad IQR = Q_3 - Q_1$

(iii) 1.2 - 1.5 × 0.2 = 0.9 lower boundary

1.4 + 1.5 × 0.2 = 1.7 upper boundary

0.4 m is an outlier

The heights, in metres, of a flock of 20 flamingos are recorded and shown below:

0.4	0.9	1.0	1.0	1.2	1.2	1.2	1.2	1.2	1.2
1.3	1.3	1.3	1.4	1.4	1.4	1.4	1.5	1.5	1.6

An outlier is an observation that falls either more than 1.5 × (interquartile range) above the upper quartile or less than 1.5 × (interquartile range) below the lower quartile.

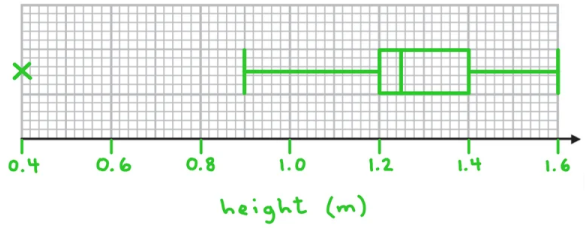
(a) (i) Find the values of Q_1, Q_2 and Q_3.

(ii) Find the interquartile range.

(iii) Identify any outliers.

[4]

(b) Using your answers to part (a), draw a box plot for the data.



[3]

b)

a) (i) Median is $\frac{20+1}{2} = 10.5^{th}$ value:

$Q_2 = \frac{1.2+1.3}{2} = 1.25$ m

$\frac{20}{4} = 5 \Rightarrow Q_1$ is 5.5^{th} value

$20 \times \frac{3}{4} = 15 \Rightarrow Q_3$ is 15.5^{th} value

$Q_1 = \frac{1.2+1.2}{2} = 1.2$ m $\quad Q_3 = \frac{1.4+1.4}{2} = 1.4$ m

(ii) IQR = 1.4 − 1.2 = 0.2 m $\quad IQR = Q_3 - Q_1$

(iii) $1.2 - 1.5 \times 0.2 = 0.9$ lower boundary

$1.4 + 1.5 \times 0.2 = 1.7$ upper boundary

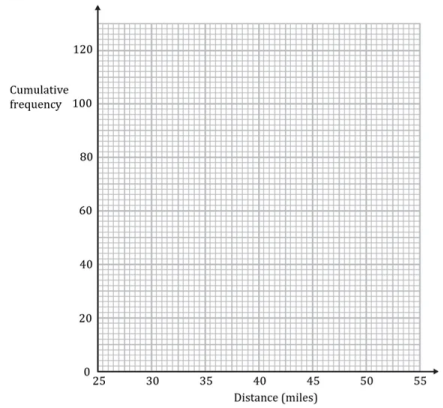
0.4 m is an outlier

Question 7

120 competitors enter an elimination race for charity. Runners set off from the same start running as many laps of the course as possible. Their total distance is tracked and the competitor who runs the furthest over a 6-hour period is the winner. The distances runners achieved are recorded in the table below:

Distance d (miles)	Frequency f	
$25 \le d < 30$	8	0
$30 \le d < 35$	10	8
$35 \le d < 40$	32	18
$40 \le d < 45$	54	50
$45 \le d < 50$	10	104
$50 \le d < 55$	6	114
		120

(a) On the grid below, draw a cumulative frequency graph for the information in the table.

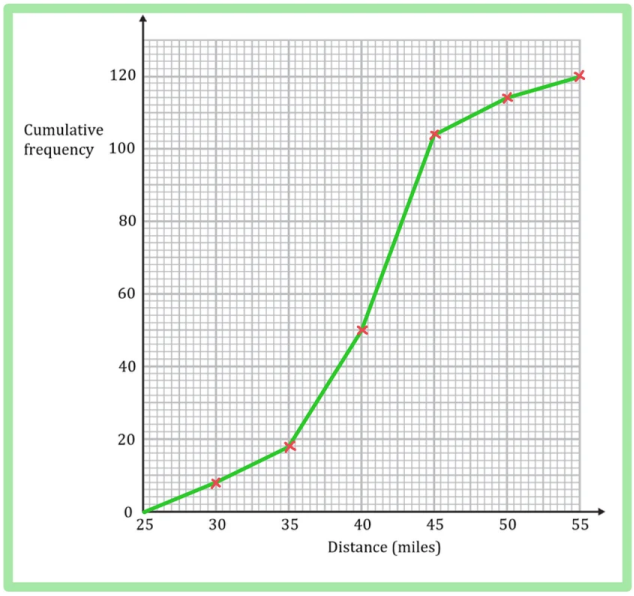


[3]

(b) Use your graph to find an estimate for the median and interquartile range.

[3]

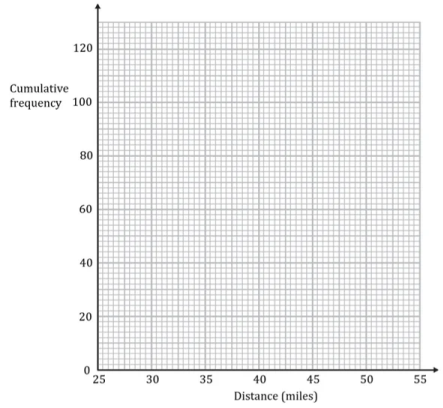
a)



120 competitors enter an elimination race for charity. Runners set off from the same start running as many laps of the course as possible. Their total distance is tracked and the competitor who runs the furthest over a 6-hour period is the winner. The distances runners achieved are recorded in the table below:

Distance d (miles)	Frequency f	
$25 \le d < 30$	8	0
$30 \le d < 35$	10	8
$35 \le d < 40$	32	18
$40 \le d < 45$	54	50
$45 \le d < 50$	10	104
$50 \le d < 55$	6	114
		120

(a) On the grid below, draw a cumulative frequency graph for the information in the table.

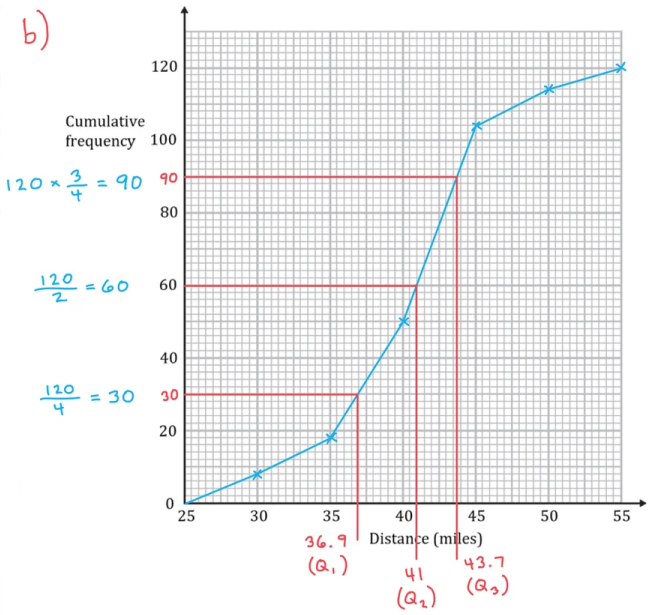


[3]

(b) Use your graph to find an estimate for the median and interquartile range.

[3]

b)

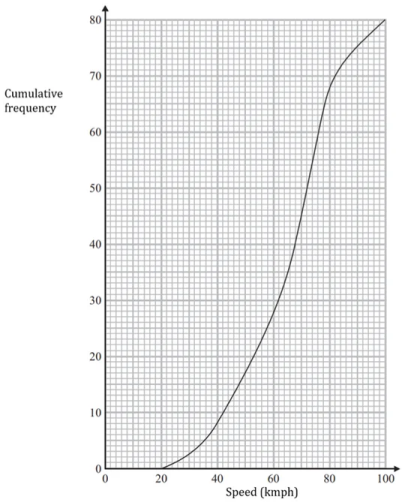


median = 41 miles

IQR = 43.7 − 36.9 = 6.8 miles

Question 8

Police check the speed of vehicles travelling along a stretch of highway. The cumulative frequency curve below summarises the data for the speeds, in kmph, of 80 vehicles:



(a) Use the graph to find an estimate for the median speed.

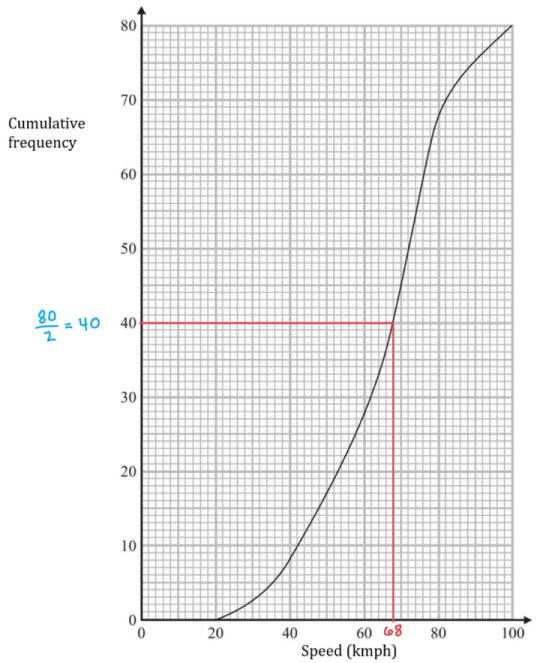
[2]

The speed limit for this section of road is 80 kmph.

(b) Vehicles travelling above the speed limit are issued with a speeding ticket. Those travelling more than 10% over the speed limit are pulled over. Use the graph to estimate the percentage of vehicles that the police pull over.

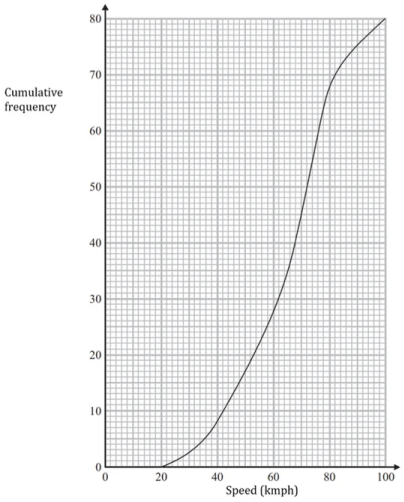
[3]

a)



median = 68 kmph

Police check the speed of vehicles travelling along a stretch of highway. The cumulative frequency curve below summarises the data for the speeds, in kmph, of 80 vehicles:



(a) Use the graph to find an estimate for the median speed.

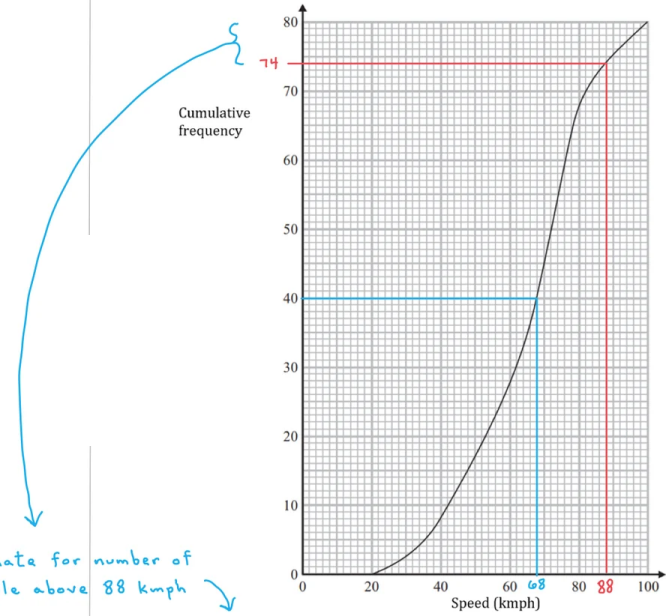
[2]

The speed limit for this section of road is 80 kmph.

(b) Vehicles travelling above the speed limit are issued with a speeding ticket. Those travelling more than 10% over the speed limit are pulled over. Use the graph to estimate the percentage of vehicles that the police pull over.

[3]

b) 10% over 80 kmph = 88 kmph



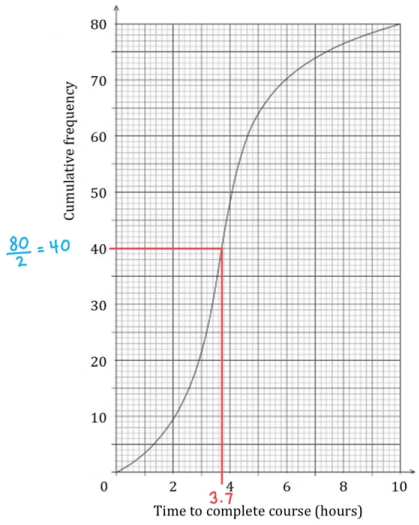
Estimate for number of people above 88 kmph

80 - 74 = 6

$\frac{6}{80} \times 100 = 7.5\%$

Question 9

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



(a) Find the median number of hours spent completing the online driving course.

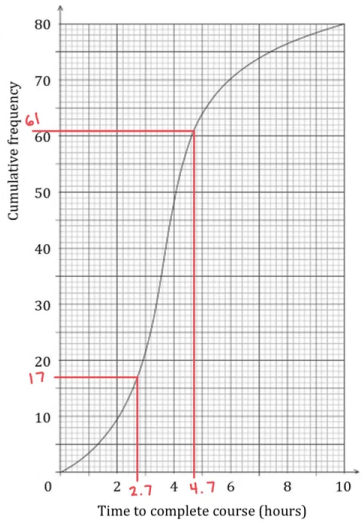
[2]

(b) Find the number of students whose online course time was within 1 hour of the median.

[2]

a) Median = 3.7 hours

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



(a) Find the median number of hours spent completing the online driving course.

Median = 3.7 hours

[2]

(b) Find the number of students whose online course time was within 1 hour of the median.

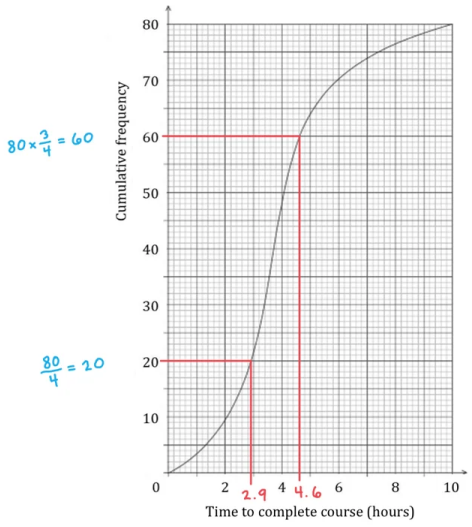
3.7 − 1 = 2.7
3.7 + 1 = 4.7

[2]

b) 61 − 17 = 44

44 students

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



(c) Calculate the interquartile range.

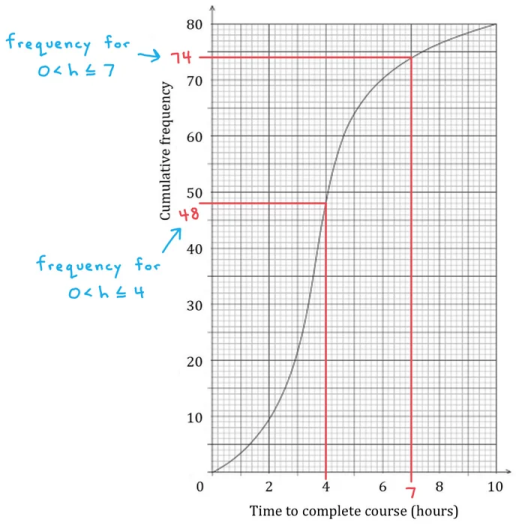
[2]

c) LQ = 2.9 UQ = 4.6

4.6 − 2.9 = 1.7 IQR = UQ − LQ

IQR = 1.7 hours

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



The same information is represented by the following table.

Hours, h	$0 < h \leq 2$	$2 < h \leq 4$	$4 < h \leq 7$	$7 < h \leq 10$
Frequency	9	p	q	6

(d) Find the value of p and the value of q.

[3]

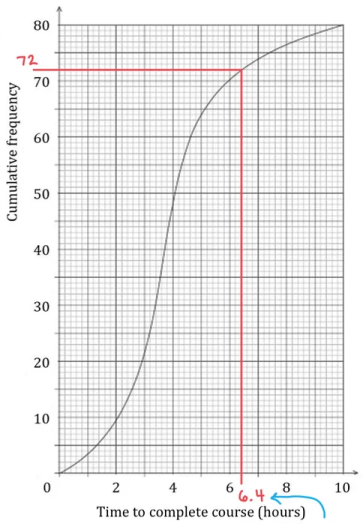
d) $p = 48 - 9 = 39$

$q = 74 - 48 = 26$

$p = 39$ $q = 26$

$9 + 39 + 26 + 6 = 80$ ✓

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



It is known that 10% of students take longer than d hours to complete the online driving course.

(e) Find the value of d.

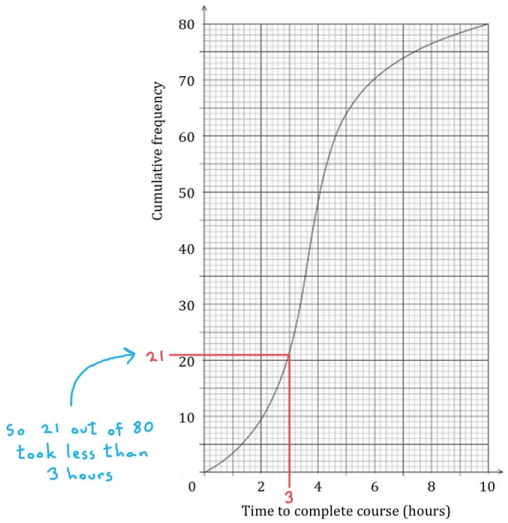
[3]

e) 10% of 80 = 8

80 − 8 = 72

$d = 6.4$

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



It is known that over a 12 month period, 4000 students in total sat the online driving course.

(f) Estimate the number of students over a 12 month period who took less than 3 hours to complete the course.

[3]

F) $4000 \times \frac{21}{80} = 1050$

Approximately 1050 students took less than 3 hours.