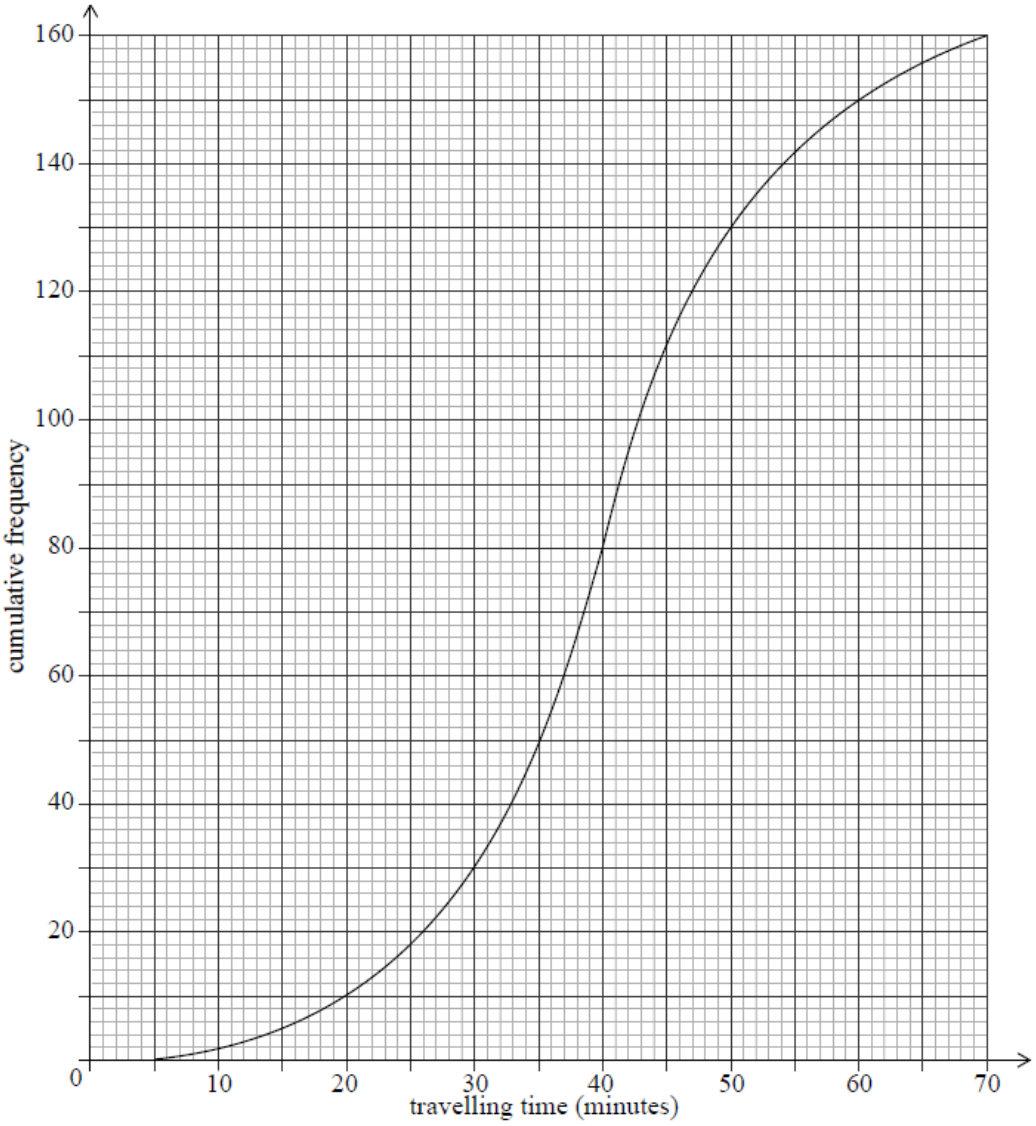
Stats - review (TL) [97 marks]

1. [Maximum mark: 15] SPM.1.SL.TZ0.7

A large company surveyed 160 of its employees to find out how much time they spend traveling to work on a given day. The results of the survey are shown in the following cumulative frequency diagram.



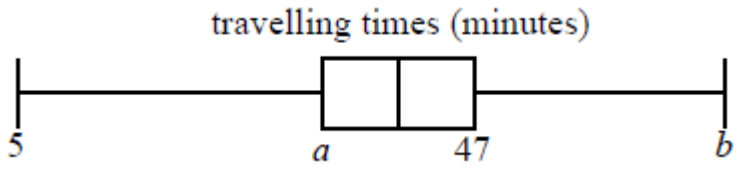
(a) Find the median number of minutes spent traveling to work. [2]

(b) Find the number of employees whose travelling time is within 15 minutes of the median. [3]

Only 10% of the employees spent more than k minutes traveling to work.

(c) Find the value of k. [3]

The results of the survey can also be displayed on the following box-and-whisker diagram.



(d) Write down the value of b. [1]

(e.i) Find the value of a. [2]

(e.ii) Hence, find the interquartile range. [2]

(f) Travelling times of less than p minutes are considered outliers.

Find the value of p. [2]

2. [Maximum mark: 4] EXN.2.SL.TZ0.1

A data set consisting of 16 test scores has mean 14.5. One test score of 9 requires a second marking and is removed from the data set.

Find the mean of the remaining 15 test scores. [4]

3. [Maximum mark: 8] EXN.2.SL.TZ0.4

The following table shows the systolic blood pressures, p mmHg, and the ages, t years, of 6 male patients at a medical clinic.

Patient	P1	P2	P3	P4	P5	P6
t (years)	40	72	35	47	21	61
p (mmHg)	105	145	100	130	95	132

(a.i) Determine the value of Pearson's product-moment correlation coefficient, r, for these data. [2]

(a.ii) Interpret, in context, the value of r found in part (a) (i). [1]

The relationship between t and p can be modelled by the regression line of p on t with equation $p = at + b$.

(b) Find the equation of the regression line of p on t. [2]

A 50-year-old male patient enters the medical clinic for his appointment.

(c) Use the regression equation from part (b) to predict this patient's systolic blood pressure. [2]

(d) A 16-year-old male patient enters the medical clinic for his appointment.

Explain why the regression equation from part (b) should not be used to predict this patient's systolic blood pressure. [1]

4. [Maximum mark: 5] 22N.2.SL.TZ0.1

The following table shows the Mathematics test scores (x) and the Science test scores (y) for a group of eight students.

Mathematics scores (x)	64	68	72	75	80	82	85	86
Science scores (y)	67	72	77	76	84	83	89	91

The regression line of y on x for this data can be written in the form $y = ax + b$.

(a) Find the value of a and the value of b. [2]

(b) Write down the value of the Pearson's product-moment correlation coefficient, r. [1]

(c) Use the equation of your regression line to predict the Science test score for a student who has a score of 78 on the Mathematics test. Express your answer to the nearest integer. [2]

5. [Maximum mark: 4] 22M.2.SL.TZ1.2

The number of hours spent exercising each week by a group of students is shown in the following table.

Exercising time (in hours)	Number of students
2	5
3	1
4	4
5	3
6	x

The median is 4.5 hours.

(a) Find the value of x. [2]

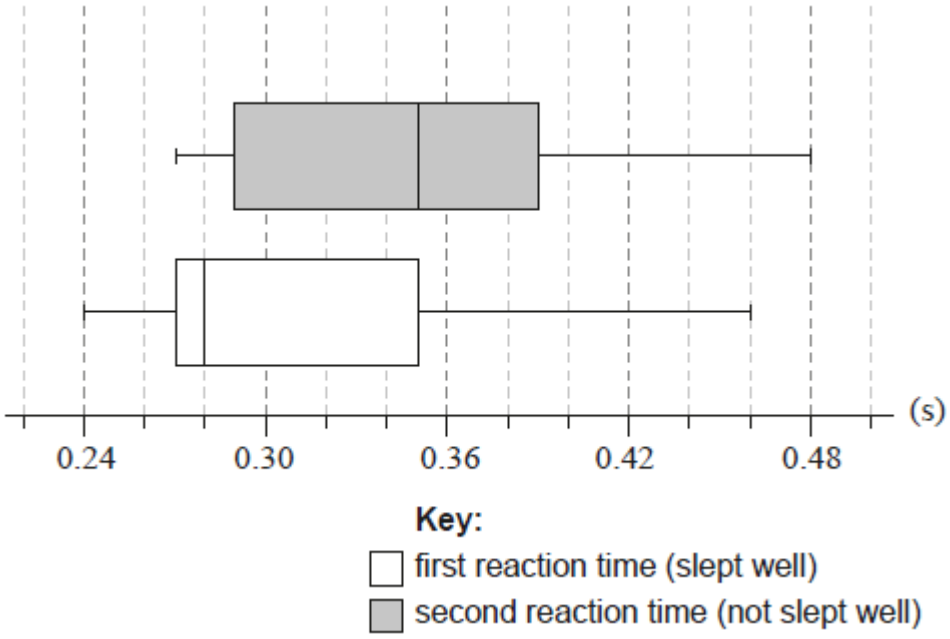
(b) Find the standard deviation. [2]

6. [Maximum mark: 6] 22M.2.SL.TZ2.5

A random sample of nine adults were selected to see whether sleeping well affected their reaction times to a visual stimulus. Each adult's reaction time was measured twice.

The first measurement for reaction time was taken on a morning after the adult had slept well. The second measurement was taken on a morning after the same adult had not slept well.

The box and whisker diagrams for the reaction times, measured in seconds, are shown below.



Consider the box and whisker diagram representing the reaction times after sleeping well.

(a) State the median reaction time after sleeping well. [1]

(b) Verify that the measurement of 0.46 seconds is not an outlier. [3]

(c) State why it appears that the mean reaction time is greater than the median reaction time. [1]

(d) Now consider the two box and whisker diagrams.

Comment on whether these box and whisker diagrams provide any evidence that might suggest that not sleeping well causes an increase in reaction time. [1]

7. [Maximum mark: 5] 21N.2.SL.TZ0.1

In Lucy's music academy, eight students took their piano diploma examination and achieved scores out of 150. For her records, Lucy decided to record the average number of hours per week each student reported practising in the weeks prior to their examination. These results are summarized in the table below.

Average weekly practice time (h)	28	13	45	33	17	29	39	36
Diploma score (D)	115	82	120	116	79	101	110	121

(a) Find Pearson's product-moment correlation coefficient, r, for these data. [2]

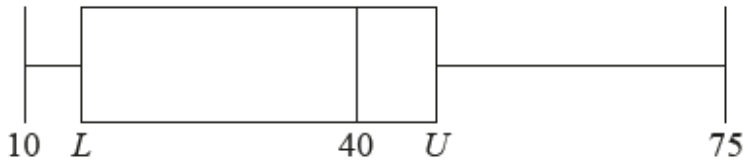
(b) The relationship between the variables can be modelled by the regression equation $D = ah + b$. Write down the value of a and the value of b. [1]

(c) One of these eight students was disappointed with her result and wished she had practised more. Based on the given data, determine how her score could have been expected to alter had she practised an extra five hours per week. [2]

8. [Maximum mark: 5] 21M.1.SL.TZ1.4

A research student weighed lizard eggs in grams and recorded the results. The following box and whisker diagram shows a summary of the results where L and U are the lower and upper quartiles respectively.

diagram not to scale



The interquartile range is 20 grams and there are no outliers in the results.

(a) Find the minimum possible value of U. [3]

(b) Hence, find the minimum possible value of L. [2]

9. [Maximum mark: 6] 21M.2.SL.TZ2.1

At a café, the waiting time between ordering and receiving a cup of coffee is dependent upon the number of customers who have already ordered their coffee and are waiting to receive it.

Sarah, a regular customer, visited the café on five consecutive days. The following table shows the number of customers, x, ahead of Sarah who have already ordered and are waiting to receive their coffee and Sarah's waiting time, y minutes.

Number of customers (x)	3	9	11	10	5
Sarah's waiting time (y)	6	10	12	11	6

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a.i) Find the value of a and the value of b. [2]

(a.ii) Write down the value of Pearson's product-moment correlation coefficient, r. [1]

(b) Interpret, in context, the value of a found in part (a)(i). [1]

(c) On another day, Sarah visits the café to order a coffee. Seven customers have already ordered their coffee and are waiting to receive it.

Use the result from part (a)(i) to estimate Sarah's waiting time to receive her coffee. [2]

10. [Maximum mark: 6] 20N.1.SL.TZ0.T_3

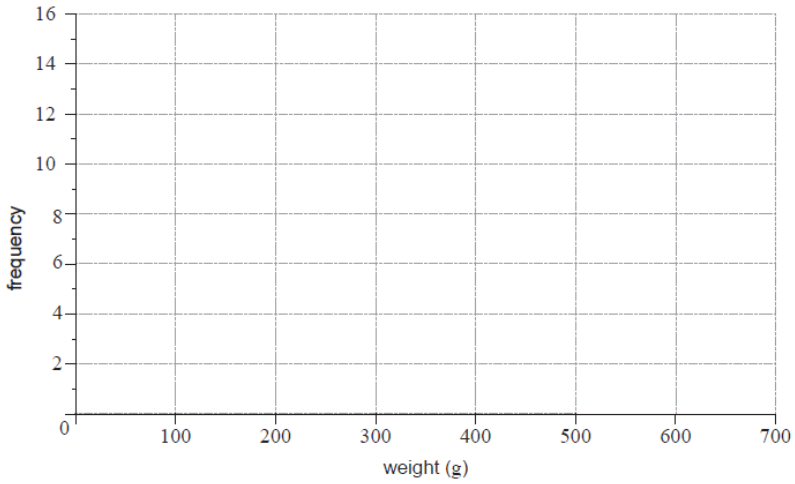
Hafizah harvested 49 mangoes from her farm. The weights of the mangoes, w, in grams, are shown in the following grouped frequency table.

Weight (g)	$100 \leq w < 200$	$200 \leq w < 300$	$300 \leq w < 400$	$400 \leq w < 500$	$500 \leq w < 600$
Frequency	4	7	14	16	8

(a) Write down the modal group for these data. [1]

(b) Use your graphic display calculator to find an estimate of the standard deviation of the weights of mangoes from this harvest. [2]

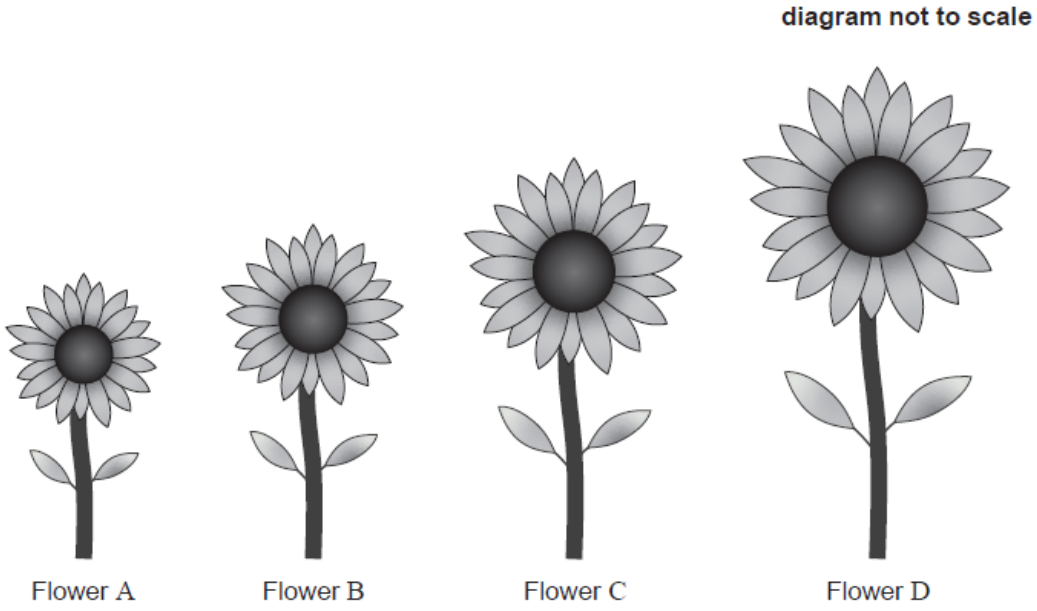
(c) On the grid below, draw a histogram for the data in the table.



[3]

11. [Maximum mark: 6] 20N.1.SL.TZ0.T_7

Anne-Marie planted four sunflowers in order of height, from shortest to tallest.



Flower C is $32\,\mathrm{cm}$ tall.

The median height of the flowers is $24\,\mathrm{cm}$.

(a) Find the height of Flower null. [2]

The range of the heights is $50\,\mathrm{cm}$. The height of Flower A is $p\,\mathrm{cm}$ and the height of Flower D is $q\,\mathrm{cm}$.

(b) Using this information, write down an equation in p and q. [1]

The mean height of the flowers is $27\,\mathrm{cm}$.

(c) Write down a second equation in p and q. [1]

(d.i) Using your answers to **parts (b)** and **(c)**, find the height of Flower A. [1]

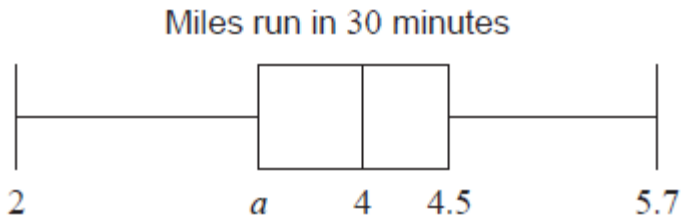
(d.ii) Using your answers to **parts (b)** and **(c)**, find the height of Flower D.

12. [Maximum mark: 15] 20N.1.SL.TZ0.S_8

Each athlete on a running team recorded the distance (M miles) they ran in 30 minutes.

The median distance is 4 miles and the interquartile range is 1.1 miles.

This information is shown in the following box-and-whisker plot.



(a) Find the value of a. [2]

The distance in miles, M, can be converted to the distance in kilometres, K, using the formula $K = \frac{8}{5}M$.

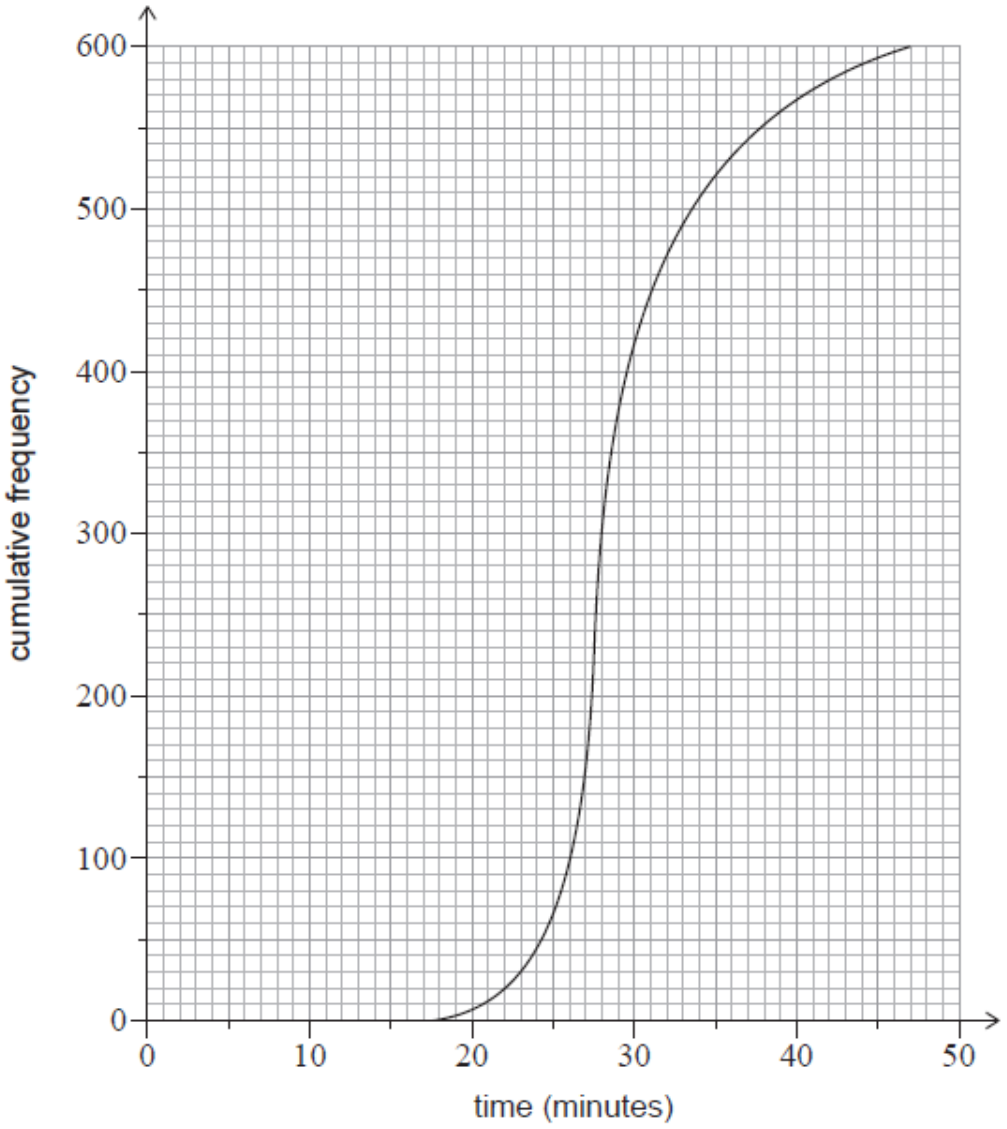
(b) Write down the value of the median distance in kilometres (km). [1]

The variance of the distances run by the athletes is $\frac{16}{9}\,\text{km}^2$.

The standard deviation of the distances is b miles.

(c) Find the value of b. [4]

A total of 600 athletes from different teams compete in a $5\,\text{km}$ race. The times the 600 athletes took to run the $5\,\text{km}$ race are shown in the following cumulative frequency graph.



There were 400 athletes who took between 22 and m minutes to complete the $5\,\text{km}$ race.

(d) Find m. [3]

(e) The first 150 athletes that completed the race won a prize.

Given that an athlete took between 22 and m minutes to complete the $5\,\text{km}$ race, calculate the probability that they won a prize. [5]

13. [Maximum mark: 6] 19M.1.SL.TZ1.T_2

The fastest recorded speeds of eight animals are shown in the following table.

Animal	Speed (km h^{-1})
Golden eagle	300
Swordfish	97
Hare	80
Lion	80
Horse	71
Zebra	64
Komodo dragon	21
Tiger beetle	6

(a) State whether **speed** is a continuous or discrete variable. [1]

(b) Write down the median speed for these animals. [1]

(c) Write down the range of the animal speeds. [1]

(d.i) For these eight animals find the mean speed. [2]

(d.ii) For these eight animals write down the standard deviation. [1]

14. [Maximum mark: 6] 19M.1.SL.TZ2.T_12

University students were surveyed and asked how many hours, h, they worked each month. The results are shown in the following table.

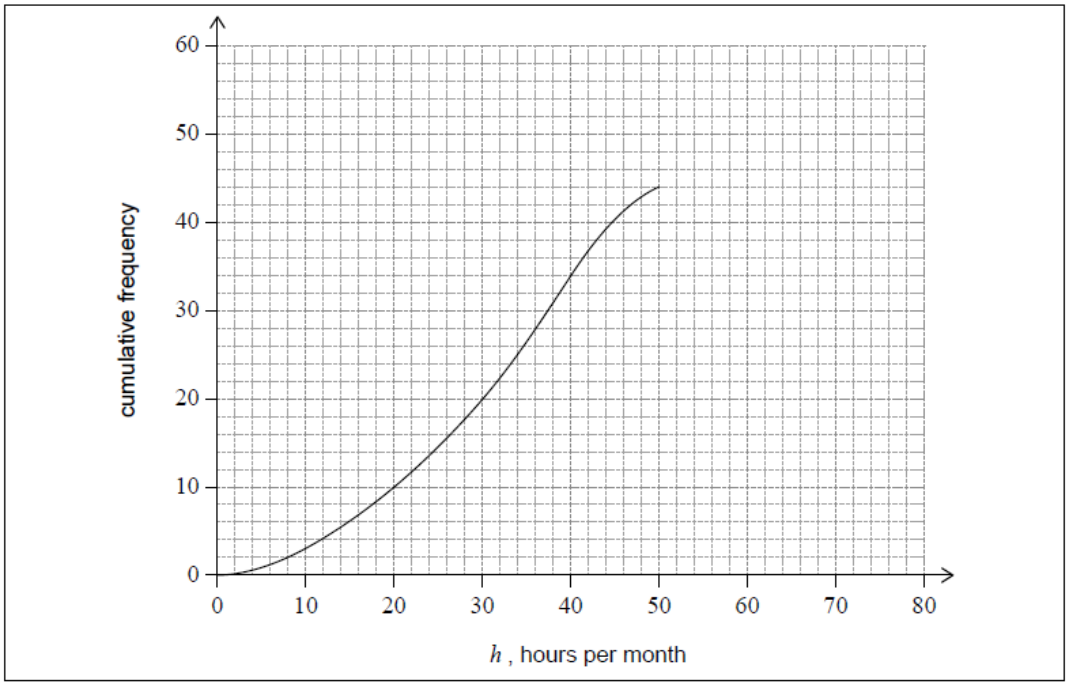
Hours per month, h	Frequency	Cumulative frequency
$0 < h \leq 10$	3	3
$10 < h \leq 20$	7	10
$20 < h \leq 30$	10	20
$30 < h \leq 40$	14	34
$40 < h \leq 50$	p	44
$50 < h \leq 60$	6	50
$60 < h \leq 70$	4	54
$70 < h \leq 80$	2	q

Use the table to find the following values.

(a.i) p. [1]

(a.ii) q. [1]

The first five class intervals, indicated in the table, have been used to draw part of a cumulative frequency curve as shown.



(b) On the same grid, complete the cumulative frequency curve for these data. [2]

(c) Use the cumulative frequency curve to find an estimate for the number of students who worked at most 35 hours per month. [2]