



General Certificate of Secondary Education
2022

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

Statistics

Unit 1
Higher Tier



MV18

[GST12]

TUESDAY 14 JUNE, AFTERNOON

Time

2 hours, plus your additional time allowance.

Instructions to Candidates

Write your Centre Number and Candidate Number in the spaces provided at the top of this page.

Write your answers in the spaces provided in this question paper.

Answer **all eleven** questions.

Any working should be clearly shown in the spaces provided since marks may be awarded for partially correct solutions.

You **may** use a calculator for this paper.

Information for Candidates

The total mark for this paper is 100.

Figures in brackets printed at the end of each question indicate the marks awarded to each question or part question.

You should have a calculator, ruler, compasses and protractor. The formula sheet is on page 2.

Higher Tier Formula Sheet

$$\text{Standard deviation} = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Spearman's Rank Correlation Coefficient

$$r_s = 1 - \left(\frac{6\sum d^2}{n(n^2 - 1)}\right)$$

Blank Page
(Questions start overleaf)

Answer **all** questions

- 1 A sports analyst researched the time, in minutes, when the first goal was scored in 15 matches in each of two football leagues.

The results are shown in the back-to-back stem and leaf diagram below.

League 1					League 2				
	7	5	4	0				9	
8	7	6	2	1	0	3		6	
	8	3	1	2	1	2	5	7	9
	9	4	2	3	2	6		6	
		1	0	4	1	2		4	

Key: 2 | 3 means 32

Key: 3 | 2 means 32

- (a) Explain why the mode is not a suitable average for the League 1 times. [1 mark]

- (b) (i) Find the median time of the first goal in League 2 matches. [1 mark]

Answer _____ minutes

(ii) Calculate the interquartile range for the League 2 matches. [2 marks]

Answer _____ minutes

The data was collected by visiting all the matches and recording times using a stopwatch.

(c) Give one advantage and one disadvantage of this method for collecting the data. [1 mark for each]

Advantage

Disadvantage

2 Colin works for a market research company. He wants to collect information on how people in a town spend their money.

Colin decides to interview people rather than sending postal questionnaires.

(a) Give two reasons why interviewing people may be better than sending postal questionnaires.

[1 mark for each]

1. _____

2. _____

Colin is going to choose between three possible sample sizes for his survey:

Sample	A	B	C
Size	12	80	1000

(b) For which sample are Colin’s results likely to be more reliable? [2 marks]

Tick one box.

A

B

C

Explain your answer.

(c) Identify one difficulty Colin may encounter if he chooses Sample C. [1 mark]

One of the questions in the survey is:

“How much money do you have in the bank?”

(d) Explain why this question is not suitable. [1 mark]

Blank Page
(Questions continue overleaf)

- 3 Melanie recorded how long she spent studying each Saturday over a period of time.

The table below shows the frequencies for the number of hours Melanie spent studying.

Time, t (hours)	Frequency	Cumulative frequency
$0 \leq t < 2$	4	
$2 \leq t < 4$	5	
$4 \leq t < 6$	18	
$6 \leq t < 8$	8	
$8 \leq t < 10$	5	

- (a) Circle two words from the list below which describe Melanie's time spent studying. [2 marks]

Categorical

Response

Quantitative

Qualitative

Continuous

Discrete

- (b) Complete the cumulative frequency column in the table above. [2 marks]

(c) How many Saturdays are represented in Melanie's data? [1 mark]

Answer _____

(d) From the data in the table, write down:

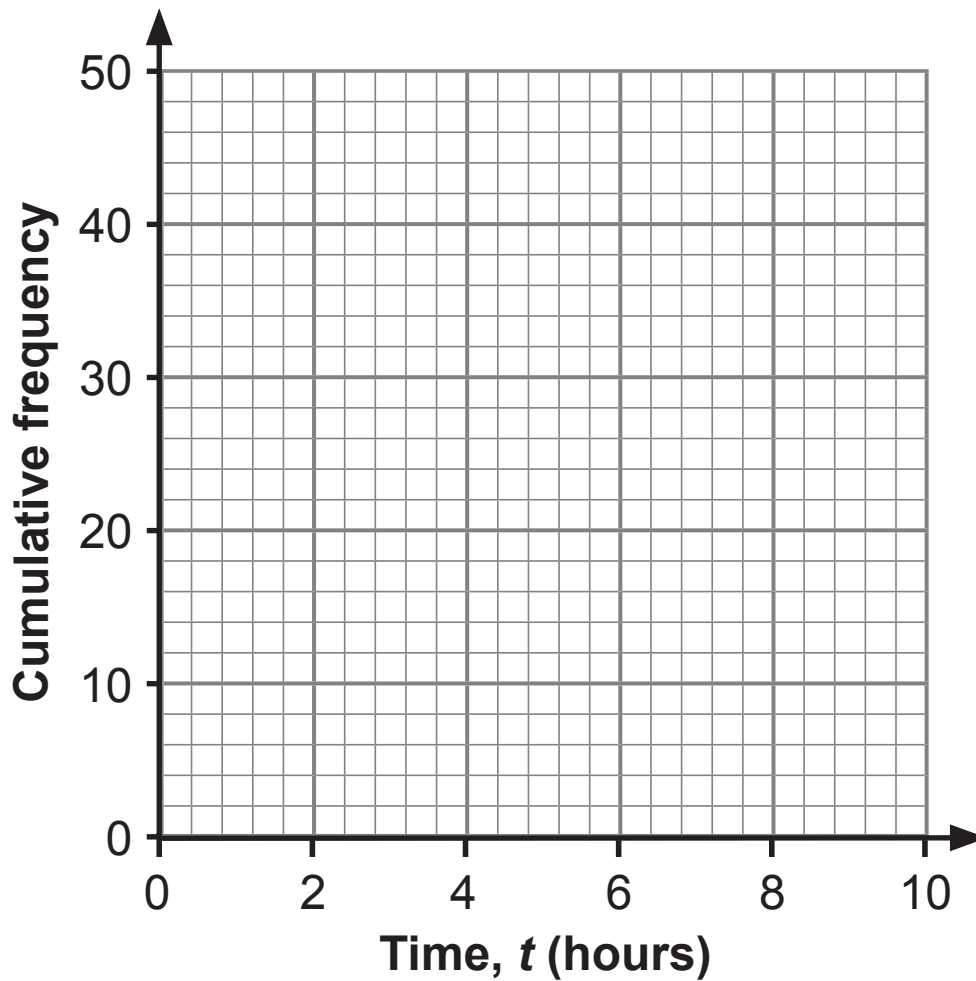
(i) the modal class; [1 mark]

Answer _____ hours

(ii) the median class. [1 mark]

Answer _____ hours

(e) On the axes below, draw a cumulative frequency diagram for the data in the table on page 10. [3 marks]



Melanie chose to use a cumulative frequency diagram rather than a histogram to display her data.

(f) Give two advantages to Melanie of using a cumulative frequency diagram to display her data rather than a histogram. [1 mark for each]

1. _____

2. _____

(g) Give one advantage to Melanie of using a histogram to display her data rather than a cumulative frequency diagram. [1 mark]

- 4 The table below shows some of the prices, in pence, and index numbers for one litre of diesel in Northern Ireland on 1 January between 2016 and 2020

The base year is 2016

Year	2016	2017	2018	2019	2020
Price (p)	114.6	101.7	121.0	122.9	
Index number	100	88.7	105.6		112.5

The index number for 2018 is 105.6

- (a) Explain what is meant by this value. [3 marks]

- (b) Calculate the index number for 2019 [3 marks]

Answer _____

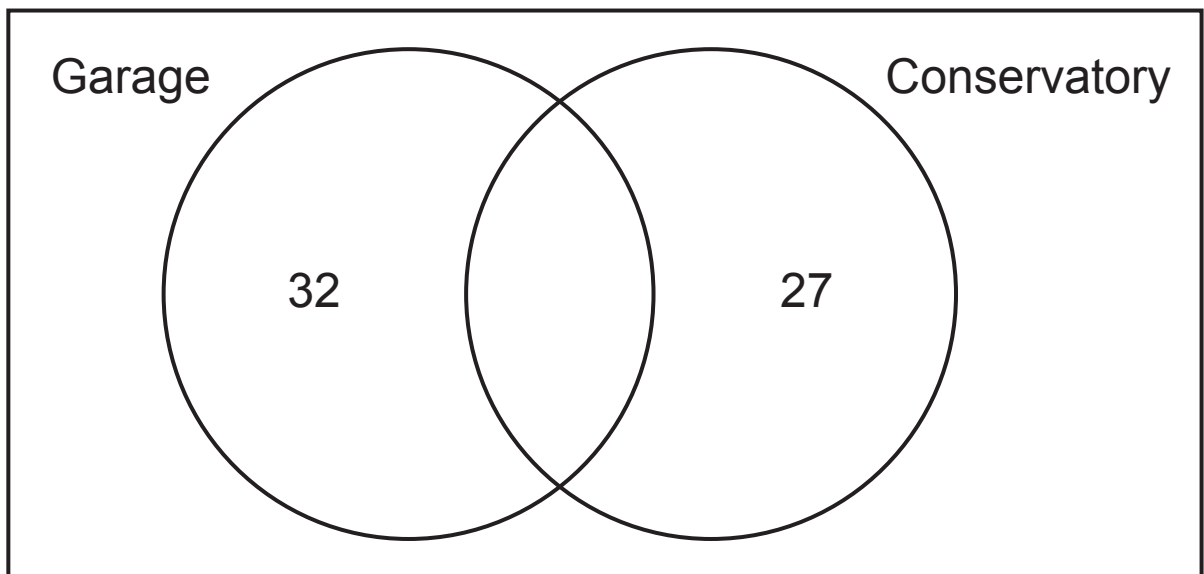
(c) Find the price of one litre of diesel on 1 January 2020
[2 marks]

Answer _____ p

- 5** An estate agent studied 85 properties to investigate how many had a garage and how many had a conservatory.

The study showed that 11 properties had both a garage and a conservatory.

- (a)** Use the information provided to complete the Venn diagram below. [2 marks]



- (b)** One property is chosen at random from the study.

Find the probability that this property:

- (i)** has a garage but not a conservatory; [1 mark]

Answer _____

(ii) does not have a garage. [2 marks]

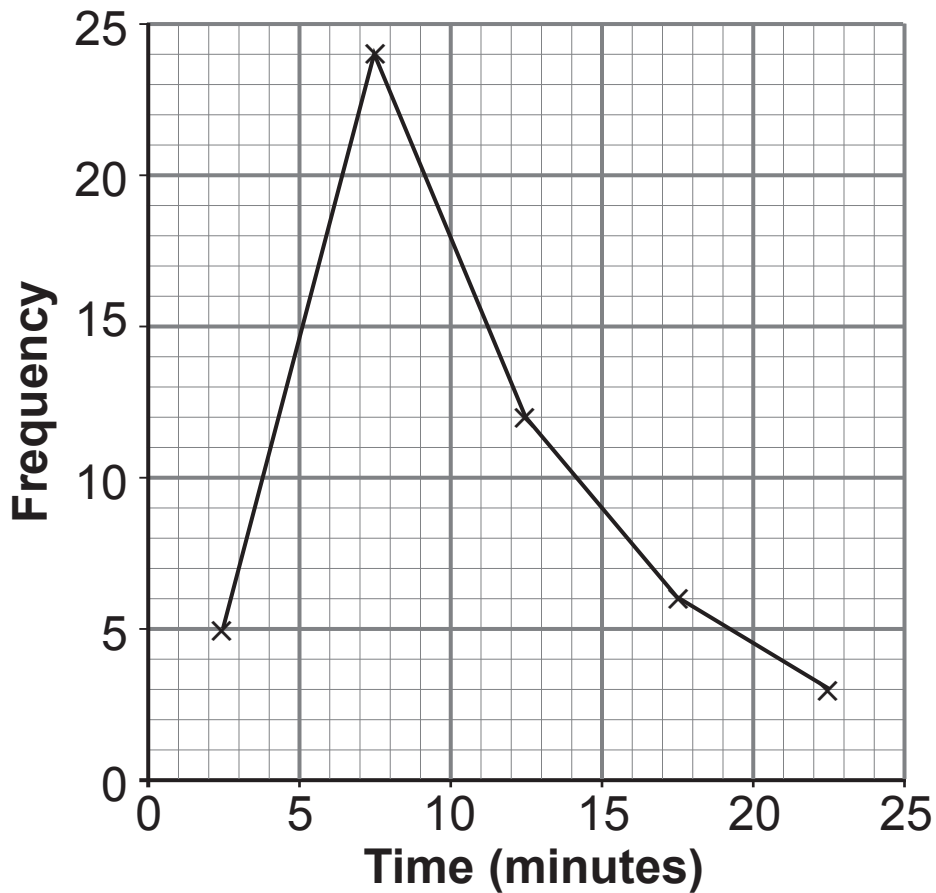
Answer _____

One property which has a conservatory is chosen at random from the study.

(c) Find the probability that this property also has a garage.
[2 marks]

Answer _____

- 6 The frequency polygon below shows the times spent by 50 customers to arrange home insurance online.



The table below shows the times taken by the same 50 customers to arrange car insurance online.

Time (mins)	$0 < x \leq 5$	$5 < x \leq 10$	$10 < x \leq 15$	$15 < x \leq 20$	$20 < x \leq 25$
Frequency	8	11	13	10	8

- (a) On the axes above, draw a frequency polygon for the car insurance data. [3 marks]

(b) Use the information provided opposite to complete the following table: [3 marks]

	Home insurance	Car insurance
Modal class		$10 < x \leq 15$
Median class		

(c) Make one comparison, in context, between the average time taken to arrange home insurance and the average time taken to arrange car insurance. [2 marks]

(d) Martin studied the data and made the following conclusion:

“The times taken to arrange car insurance were more varied than the times taken to arrange home insurance.”

Is Martin correct?

Tick one box.

Yes

No

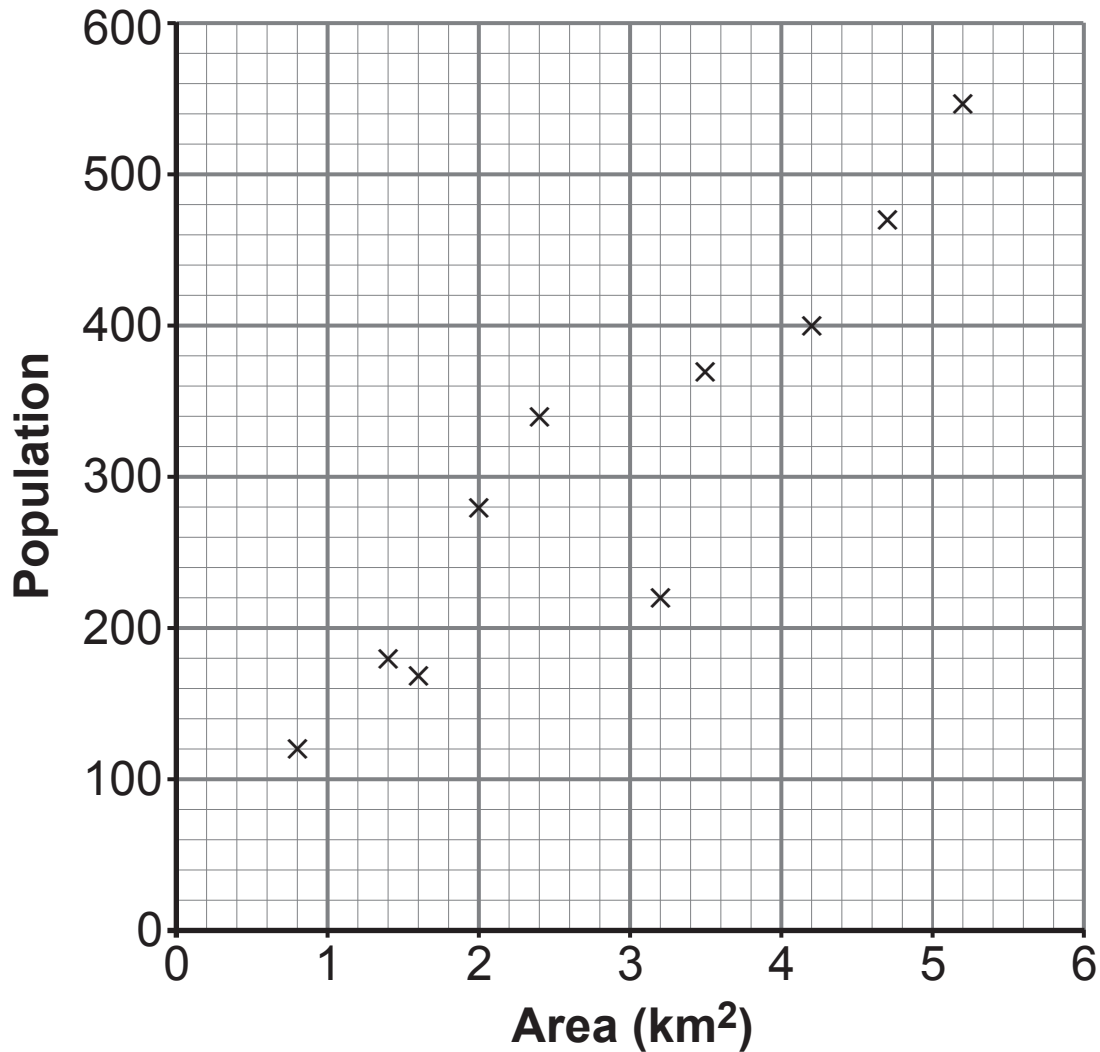
Give a reason for your answer. [2 marks]

(e) (i) Is the home insurance data positively skewed or negatively skewed? [1 mark]

(ii) Explain your choice. [1 mark]

7 Aoife uses the internet to investigate the land area and population of 10 small villages in Northern Ireland.

Her results are shown in the scatter diagram below.



(a) (i) Identify the response variable in the scatter diagram.
[1 mark]

(ii) Justify your answer. [1 mark]

The mean area of the 10 small villages was 2.9 km^2 and the mean population was 310 people.

(b) Plot the double mean point on the scatter diagram and draw a line of best fit. [2 marks]

(c) Describe and interpret the correlation shown in the scatter diagram. [2 marks]

Aoife calculates the equation of her line of best fit to be $y = 57 + 87x$

(d) Interpret, in context, the number 87 in this equation. [2 marks]

(e) Use the equation of her line of best fit to calculate an estimate of the population of a small village with a land area of 4.4 km^2 [2 marks]

Answer _____

Blank Page
(Questions continue overleaf)

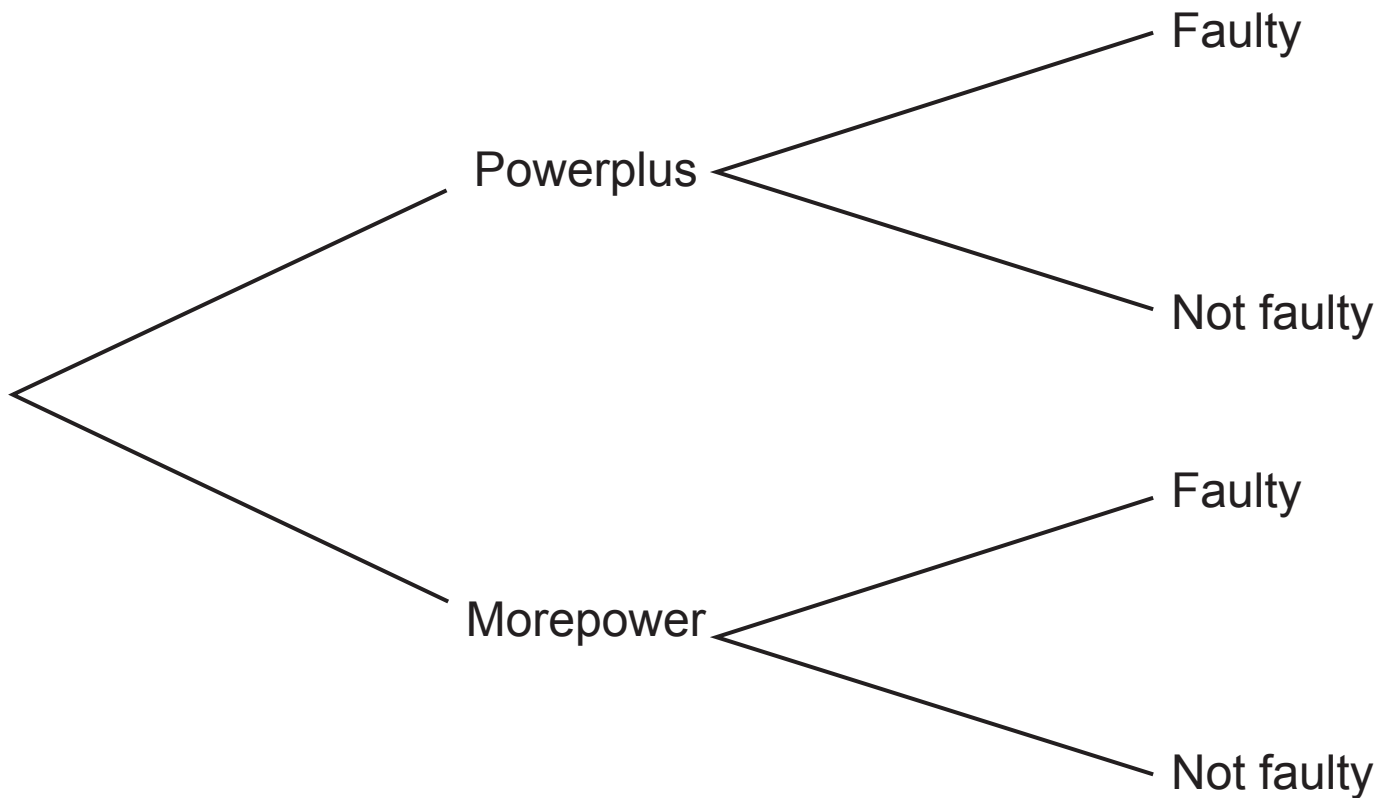
8 A factory produces two types of battery: Powerplus and Morepower.

Twice as many Powerplus batteries are produced than Morepower batteries.

It is estimated that 2% of Powerplus batteries are faulty and 3% of Morepower batteries are faulty.

Anthony selects one battery from a large mixed batch.

(a) Complete the tree diagram below. [3 marks]



(b) Calculate the relative risk of Anthony selecting a faulty Powerplus battery compared to a faulty Morepower battery. [3 marks]

Answer _____

(c) Give an interpretation of your answer to part **(b)**. [2 marks]

9 Laura works in the Human Resources department of a large company.

She recorded the number of absences every month between September 2018 and May 2019

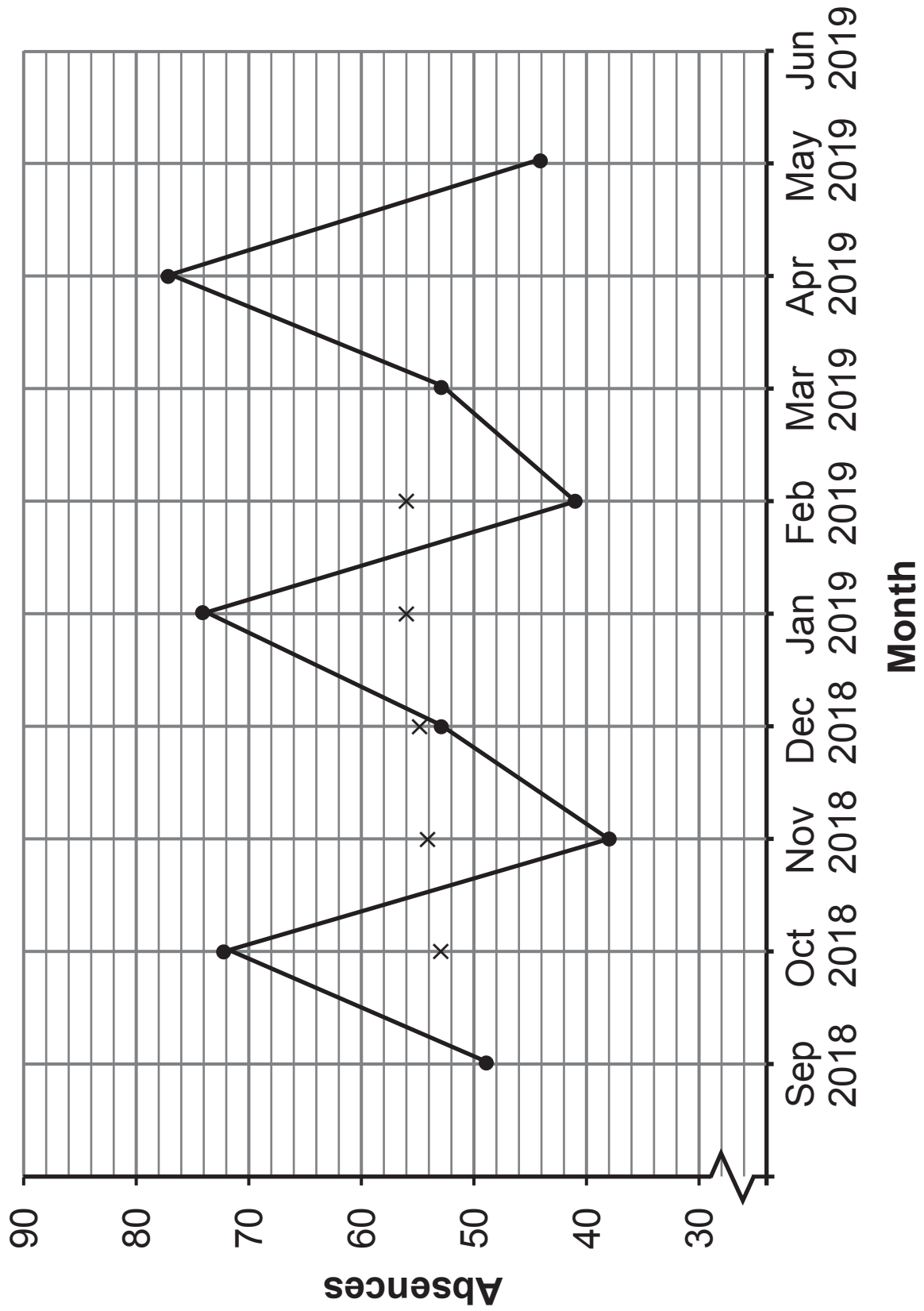
Her results are shown in the table below.

Month	Sep 2018	Oct 2018	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Apr 2019	May 2019
Absences	49	72	38	53	74	41	53	77	44

Laura's results are shown in the graph opposite.

To illustrate the trend in the data, Laura calculated 3-point moving averages.

(a) Explain why Laura calculated moving averages using 3 points. [1 mark]



Laura plotted the first five moving averages on the graph.

(b) Show that the next two moving averages are 57 and 58
[2 marks]

(c) Plot the two moving averages from part **(b)** on the graph
and draw a trend line. [3 marks]

(d) Describe the trend in the number of absences during the
nine-month period. [1 mark]

(e) Use your trend line to predict the number of absences for June 2019 [3 marks]

Answer _____

10 Sam records the typical operating temperature (in °C) of different types of engines as shown in the table below.

Temperature (°C)	Frequency			
$30 < x \leq 40$	4			
$40 < x \leq 45$	7			
$45 < x \leq 50$	8			
$50 < x \leq 60$	6			
$60 < x \leq 80$	10			

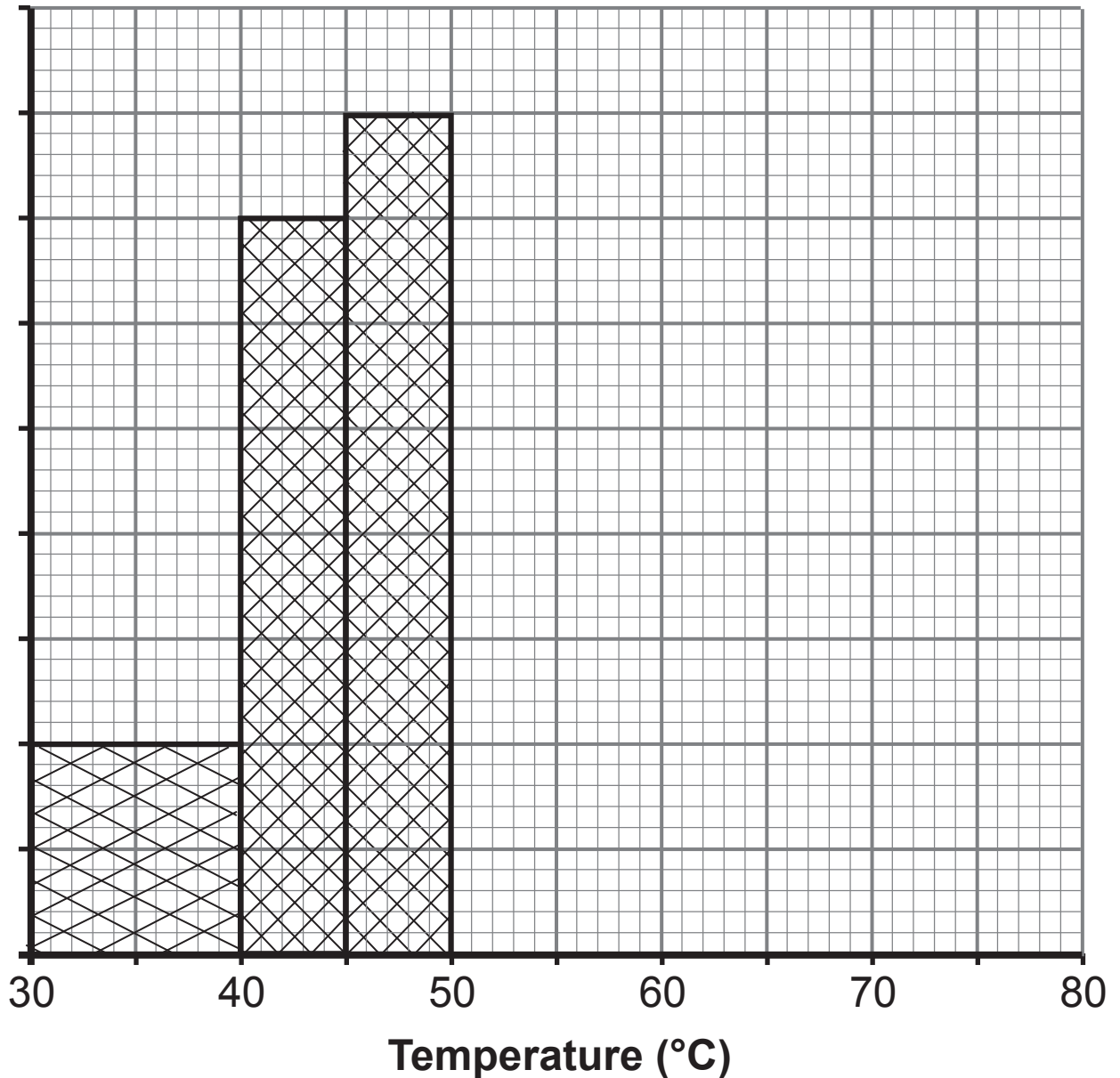
(a) Calculate an estimate of the mean temperature of the engines. [3 marks]

Give your answer correct to 3 significant figures.

Answer _____ °C

(b) Show that an estimate of the standard deviation for the temperatures is 12.2°C . [3 marks]

(c) Use the data in the table to complete the histogram below. [3 marks]



Sam decides to convert his results from degrees Celsius (C) to degrees Fahrenheit (F) using the formula:

$$F = C \times 1.8 + 32$$

(d) Using your answers to parts **(a)** and **(b)**, calculate:

- (i)** the estimated mean temperature in degrees Fahrenheit; [1 mark]

Answer _____°F

- (ii)** the estimated standard deviation in degrees Fahrenheit. [1 mark]

Answer _____°F

Blank Page
(Questions continue overleaf)

11 Harry, Kyle and Ethan entered a junior sporting competition.

The points scored by all competitors in the junior sporting competition were normally distributed with a mean of 98 and a standard deviation of 3.2

(a) Harry scored 102 points in the competition.

(i) Calculate Harry's standardised score. [2 marks]

Answer _____

Kyle's standardised score in the competition was 1.4

(ii) Who scored more points in the competition?
[2 marks]

Tick one box.

Harry

Kyle

Give a reason for your answer.

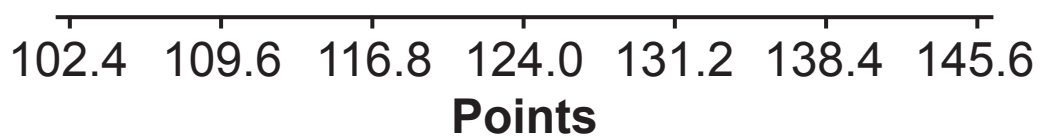
Ethan's standardised score was calculated to be -0.31

(b) Calculate how many points Ethan scored in
the competition. [2 marks]

Answer _____

In a senior sporting competition, the points scored by the 348 competitors were normally distributed with a mean of 124 and a standard deviation of 7.2

- (c) (i) Use the axis below to sketch the distribution of the points in the senior sporting competition. [1 mark]



- (ii) Calculate an estimate of the number of competitors who scored 116.8 points or less in the senior sporting competition. [2 marks]

Answer _____

This is the end of the question paper

For Examiner's use only	
Question Number	Marks
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
Total Marks	

Permission to reproduce all copyright material has been applied for. In some cases, efforts to contact copyright holders may have been unsuccessful and CCEA will be happy to rectify any omissions of acknowledgement in future if notified.