



Cambridge O Level

CANDIDATE
NAME

--

CENTRE
NUMBER

--	--	--	--	--

CANDIDATE
NUMBER

--	--	--	--



STATISTICS

4040/12

Paper 1

October/November 2022

2 hours 15 minutes

You must answer on the question paper.

You will need: Calculator
Pair of compasses
Protractor

INSTRUCTIONS

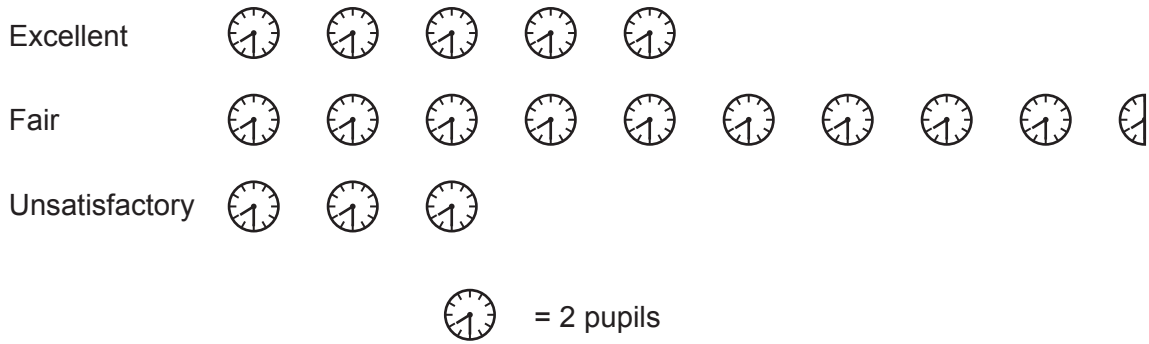
- Answer **all** questions.
- Use a black or dark blue pen. You may use an HB pencil for any diagrams or graphs.
- Write your name, centre number and candidate number in the boxes at the top of the page.
- Write your answer to each question in the space provided.
- Do **not** use an erasable pen or correction fluid.
- Do **not** write on any bar codes.
- You should use a calculator where appropriate.
- You must show all necessary working clearly.

INFORMATION

- The total mark for this paper is 100.
- The number of marks for each question or part question is shown in brackets [].

This document has **16** pages.

- 1 For her end-of-year reports a teacher categorises the punctuality of each pupil in her class as excellent, fair, or unsatisfactory. The pictogram shows the number of pupils in each category.



(a) State the number of pupils whose punctuality was excellent.
 [1]

(b) How many more pupils had fair punctuality than had unsatisfactory punctuality?
 [1]

The headteacher selects a report at random.

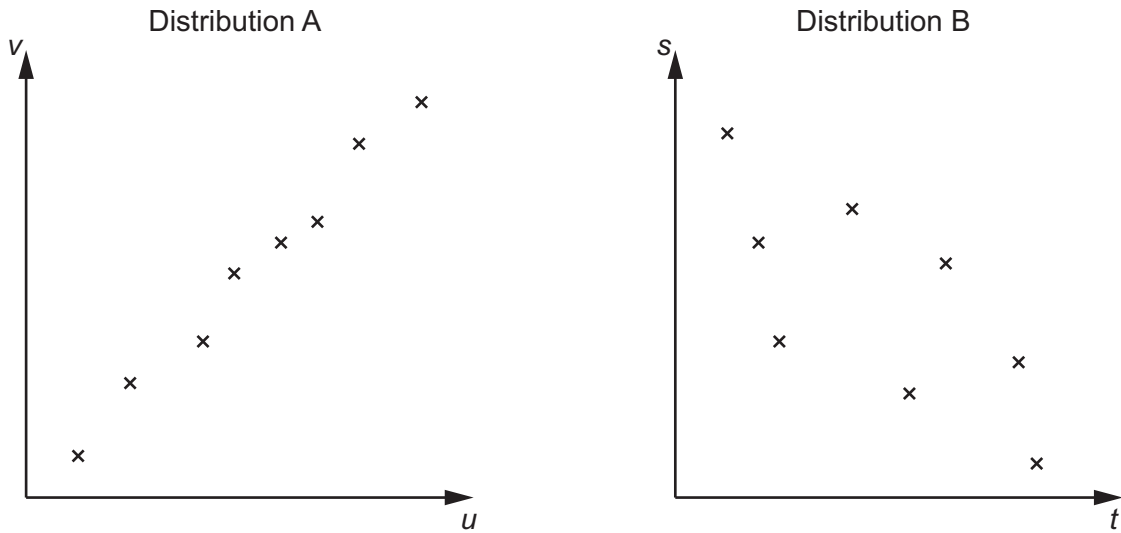
(c) Find the probability that the report shows this pupil's punctuality to have been fair or excellent.
 [1]

- 2 (a) Describe the difference between positive and negative correlation for the variables in a bivariate distribution.

.....

 [2]

- (b) The variables in two bivariate distributions were measured and plotted. The following scatter diagrams were obtained.



Describe fully the correlation shown by the variables in each of these distributions.

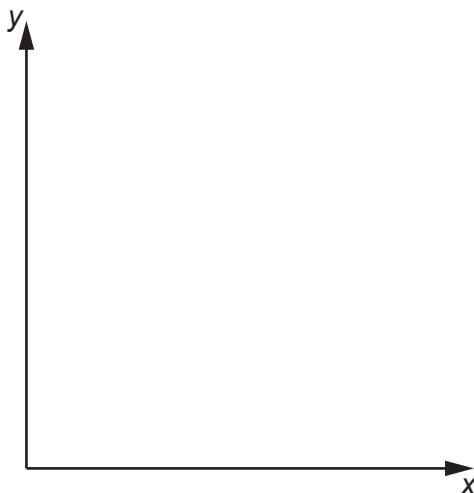
Distribution A

Distribution B

[3]

- (c) For eight schoolchildren their height, x , and the distance of their homes from school, y , were recorded.

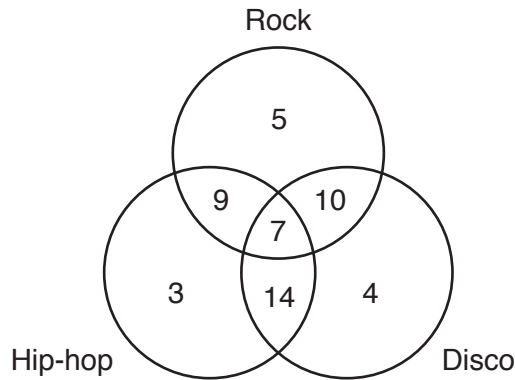
Sketch, using the axes below, a scatter diagram you might expect to obtain if the recorded values of x and y were plotted. You are **not** required to scale the axes.



[1]

[Turn over

- 3 The teenagers at a club are asked about the types of music they like. The diagram below shows the number who like one or more of Rock, Hip-hop and Disco.



- (a) Interpret the value 14 in the diagram.

.....
 [1]

- (b) Find the number of teenagers who like

(i) Disco,

..... [1]

(ii) Rock and Hip-hop,

..... [1]

(iii) only one of these three types of music.

..... [1]

- (c) Of the teenagers who like two or more of these types of music, find the percentage who like Disco and Rock.

..... [2]

- 4 Thabisa is a medical student. She conducts research into the number of visits to a dentist, and number of visits to a doctor, that people make. The table shows her results for a sample of 50 people in her town in one particular year.

		<i>Number of visits to a doctor</i>				
		0	1–3	4–6	7–9	10 or more
<i>Number of visits to a dentist</i>	0	3	5	7	1	0
	1	2	3	4	6	5
	2	1	2	3	2	1
	3–5	1	1	2	1	0

For example, there were 6 people who made 1 visit to a dentist and 7, 8 or 9 visits to a doctor.

- (a) Find the number of people who, in this year,

(i) made 2 visits to a dentist,

..... [1]

(ii) made more than 3 visits to a doctor,

..... [1]

(iii) made fewer than 2 visits to a dentist and fewer than 4 visits to a doctor.

..... [1]

- (b) Estimate the total number of visits made to a dentist by this sample of people in this year.

..... [3]

Thabisa thinks that her results show that a dentist in the town will have fewer appointments than a doctor in the town.

- (c) Explain why Thabisa might be wrong.

.....
 [1]

5 In a game, a turn consists of throwing three unbiased six-sided dice, each with faces numbered 1, 2, 3, 4, 5 and 6. The score for a turn is the sum of the two highest numbers which appear, provided no number appears more than once. If any number appears more than once the score is zero. For example, if 3, 2, 5 appear, the score is 8; if 2, 3, 2 appear, the score is zero.

(a) Write down the smallest possible non-zero score.

..... [1]

(b) Find the probability of obtaining a score of 9 in one turn.

..... [5]

(c) Find the probability of obtaining a score of 9 in each of two successive turns.

..... [2]

- 6 Darcy, Emma, George, Jane and Lizzy work in the same office. They compare the time it takes them to travel to work each morning. The table shows information on their journey times collected over a period of 10 working days.

	<i>Mean journey time</i> (minutes)	<i>Range of journey times</i> (minutes)	<i>Shortest journey time</i> (minutes)
Darcy	36.2	11	30
Emma	40.1	15	27
George	38.3	12	36
Jane	39.8	9	37
Lizzy	42.6	8	39

- (a) (i) State, with a reason, for which worker their journey times appear to have been generally least varied.

..... [1]

- (ii) State, with a reason, for which worker their journey times appear to have been generally most varied.

..... [1]

- (iii) State any assumption made in giving your answers to parts (a)(i) and (a)(ii).

.....
..... [1]

- (b) Giving your answers in minutes, find, for this period,

- (i) the total journey time for Jane,

..... [1]

- (ii) the range of the total journey times for the five workers.

..... [2]

- (c) Giving your answer in minutes, find, for this period, the range of the longest journey times of the five workers.

..... [2]

- 7 Rashid keeps a herd of goats for the milk they produce. The table summarises the daily volume of milk obtained from the herd over a period of 42 days.

<i>Daily volume of milk (litres)</i>	<i>Number of days</i>	
20–under 22	3	
22–under 24	5	
24–under 26	9	
26–under 28	14	
28–under 30	7	
30–under 32	4	

- (a) Estimate the mean and standard deviation of these volumes. Give your answers to 3 significant figures.

Mean = litres

Standard deviation = litres
[7]

Rashid considers making cheese from the milk. He estimates that he will need 12 litres of milk to make 1 kg of cheese.

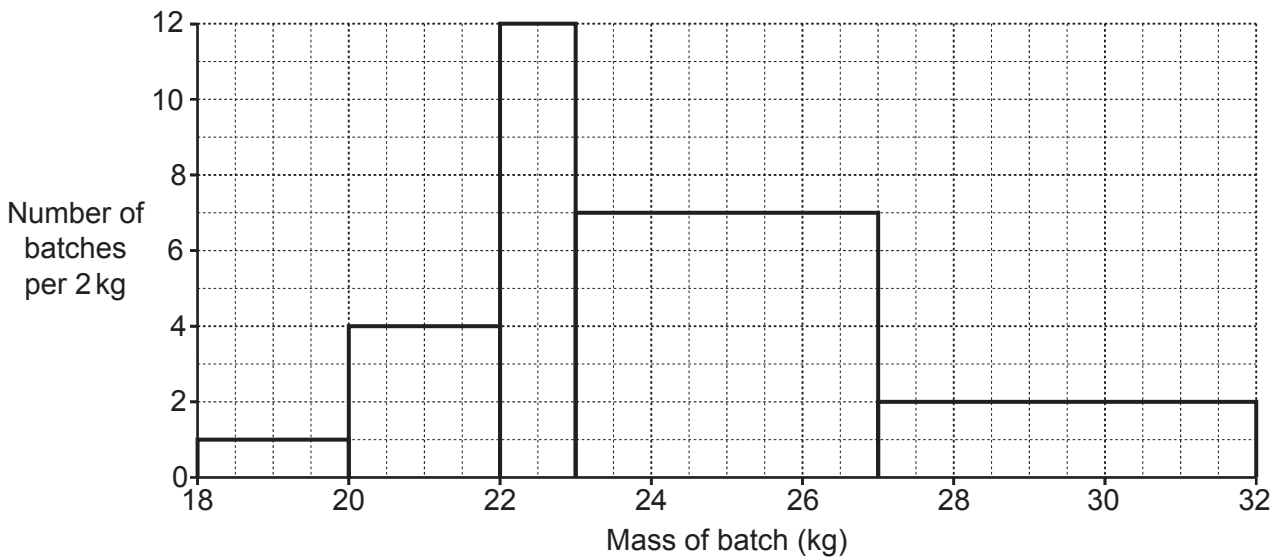
- (b) Estimate the total mass of cheese he could make from the milk produced by the herd over a period of 126 days, stating also any assumption needed in calculating your answer.

Mass of cheese kg

Assumption

..... [3]

Later, Rashid tries making yoghurt from the milk. The histogram below summarises the masses of 30 experimental batches of yoghurt.



Rashid thinks that only batches of 23 kg or more will be profitable to make.

- (c) For these batches, find the percentage which would have been profitable to make.

..... [5]

- 8 Most countries in the world experience earth tremors, even those which rarely experience earthquakes. The size of a tremor is measured by its 'magnitude' (a continuous variable). The table summarises the magnitudes of the last 80 tremors recorded at the Geological Survey Centre in a particular country.

<i>Tremor magnitude</i>	<i>Number of tremors</i>	<i>Cumulative frequency</i>
0.0–under 0.5	25	
0.5–under 1.0	18	
1.0–under 1.5	11	
1.5–under 2.0	8	
2.0–under 2.5	6	
2.5–under 3.0	5	
3.0–under 3.5	4	
3.5–under 4.0	3	

(a) Complete the cumulative frequency column in the table. [1]

(b) Plot the cumulative frequencies on the grid opposite, joining the points with a smooth curve. [3]

(c) Use the graph to estimate, for these tremor magnitudes,

(i) the median,

..... [1]

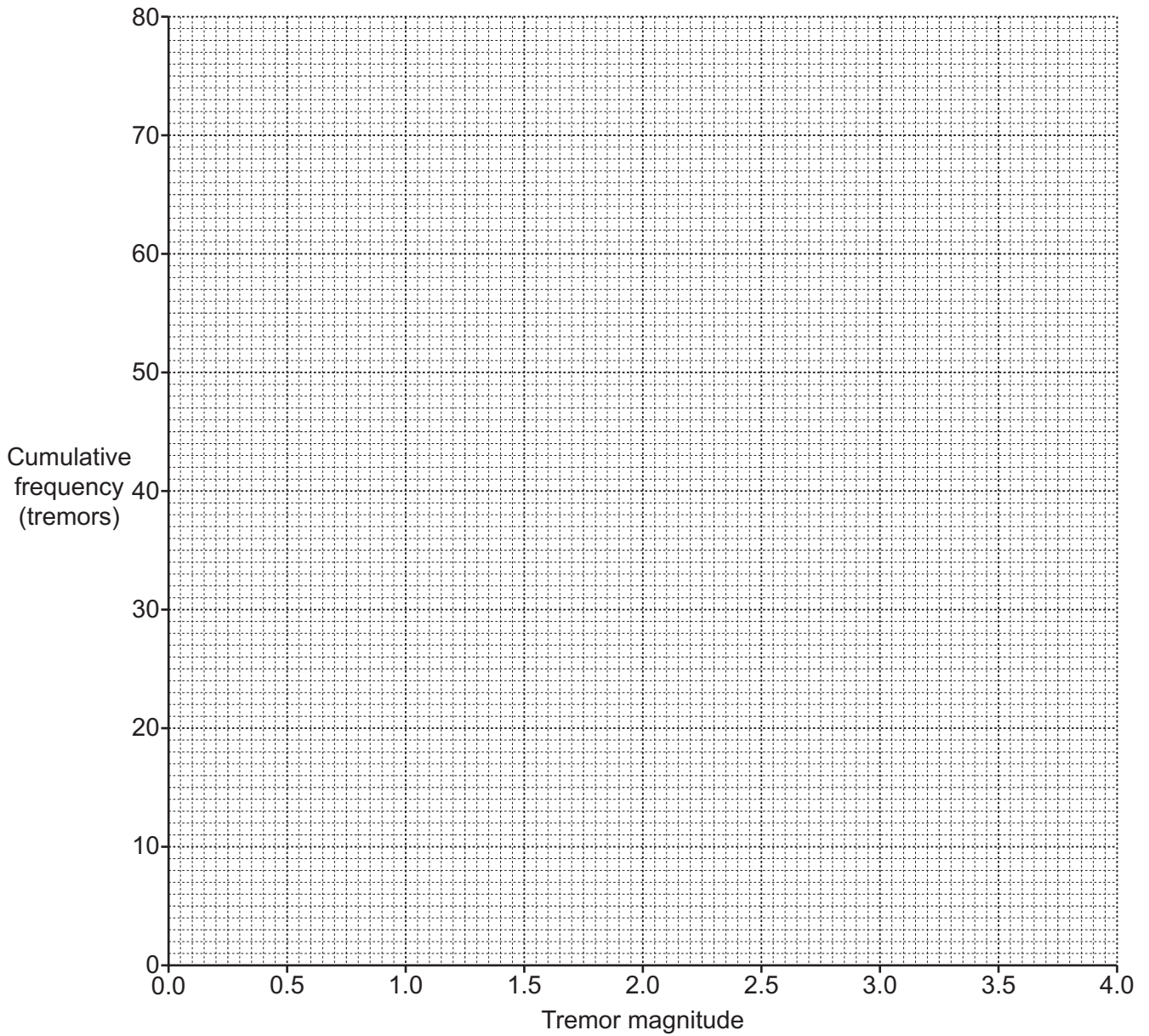
(ii) the interquartile range.

..... [4]

Many earth tremors are so small that people do not feel them. A geologist at the Survey Centre says that citizens would not have felt tremors less than 2.0 in magnitude.

(d) Use the graph to estimate the 45th percentile of magnitudes for magnitudes of 2.0 or more.

..... [2]



Two of the tremors were so large as to be classified as small earthquakes.

- (e) Use the graph to estimate the minimum magnitude for a tremor to be classified as a small earthquake.

..... [2]

- (f) Explain whether or not you think that the mean would be a better measure of central tendency than the median for this distribution. You are **not** required to carry out any further calculation.

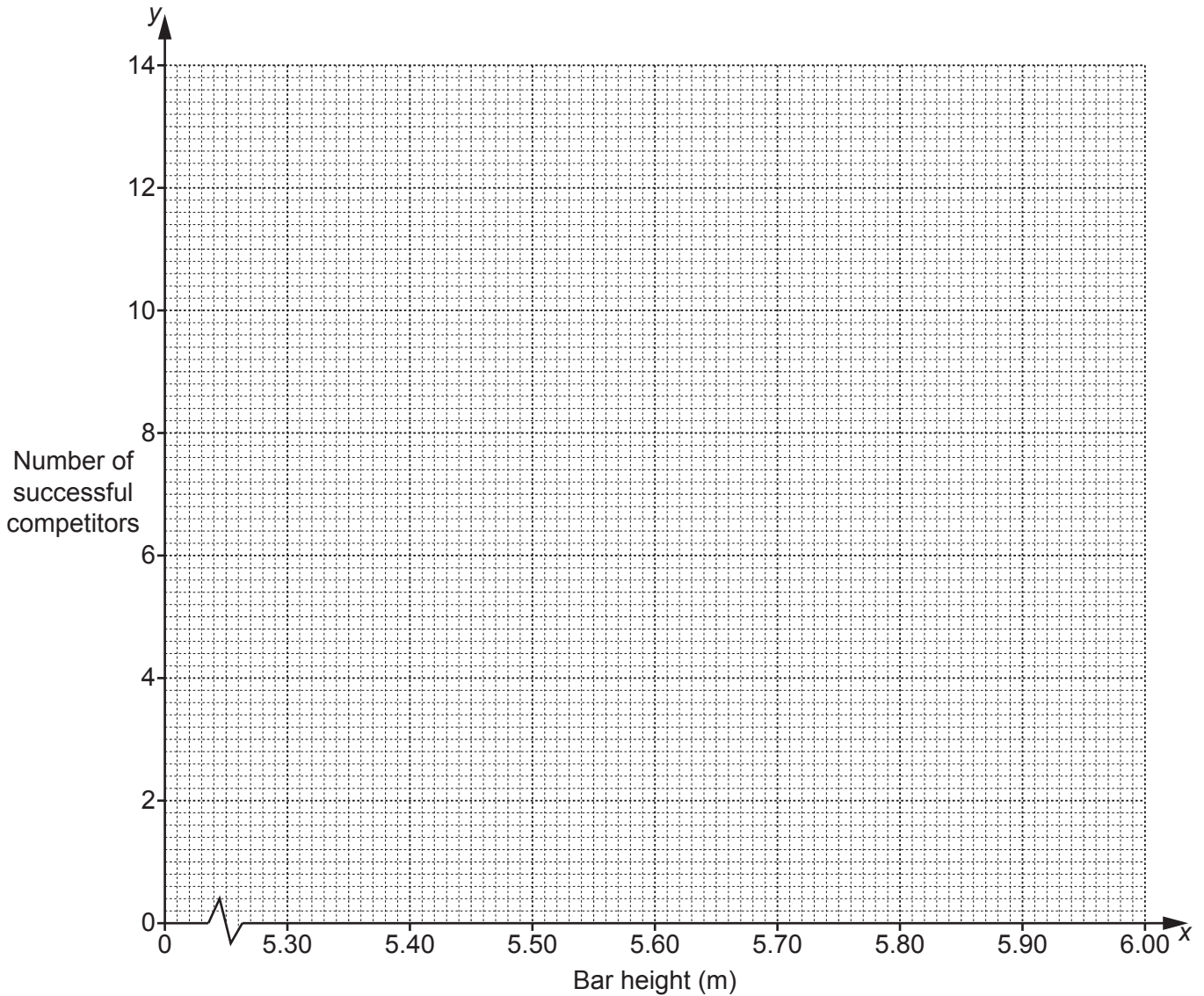
.....

 [2]

9 Natasha practises her Statistics whilst watching athletics on television. She observes that, in the pole vault, competitors attempt to clear the bar at a set height, and those who are unsuccessful are eliminated. Subsequently the bar is raised to other heights, and following further unsuccessful attempts there are further eliminations. There were 15 competitors in this event. The table shows the number who succeeded in clearing the bar at the different set heights.

Bar height, x (m)	5.30	5.40	5.50	5.60	5.70	5.75	5.80	5.85
Number of successful competitors, y	14	13	12	9	7	5	4	3

(a) Plot the data on the grid below.



[2]

(b) How many competitors had been eliminated since the start of the event following all attempts with a bar height of 5.50 m?

..... [1]

The data have an overall mean of (5.6125, 8.375) and an upper semi-average of (5.775, 4.75).

(c) Find the lower semi-average and plot this and the two given averages on your graph.

..... [3]

(d) Use your plotted averages to draw a line of best fit, and find its equation in the form $y = mx + c$.

..... [4]

(e) From the information so far collected, Natasha uses her line of best fit to estimate the height that will be cleared by the winner of the event. Use your answer to part (d) to do this yourself.

..... [2]

(f) Give one reason why Natasha should **not** have been surprised when her estimate in part (e) turned out to be inaccurate.

.....
..... [1]

After the event, the television commentator interviewed four of the competitors in the event: those finishing first and second, and two others.

(g) If the two others were chosen at random, find the probability that both had been eliminated attempting to clear a height of 5.6 m.

..... [3]

- 10 The table gives information on population numbers, and the number of deaths in road accidents, in four countries in the year 2019.

Country	Population (millions)	Number of road deaths
A	5.6	220
B	46.7	1809
C	10.3	645
D	64.1	1780

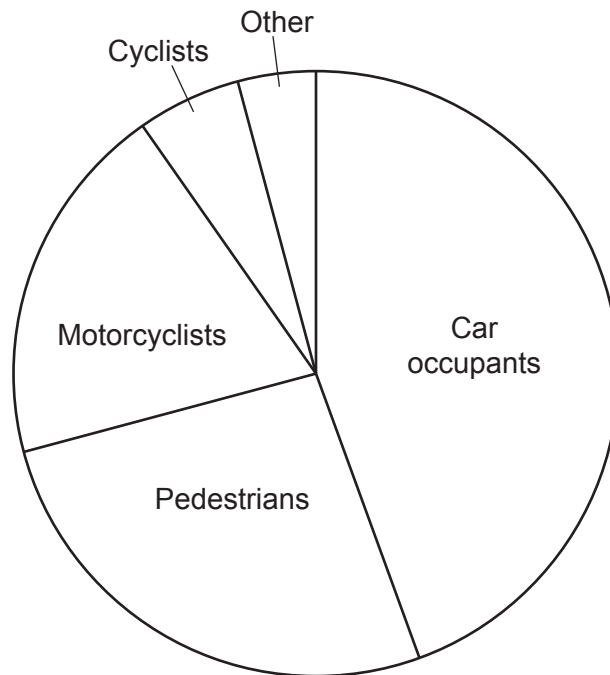
- (a) For each country calculate the crude road death rate, giving your answers per million of the population, correct to one decimal place. Hence state the country which had the least favourable road death statistics.

A B C D

Least favourable

[3]

In country D, the 1780 road deaths were categorised by road user type: Car occupants, Pedestrians, Motorcyclists, Cyclists and Other. This information is shown in the pie chart.



(b) Use the chart to find the number of these road deaths suffered by

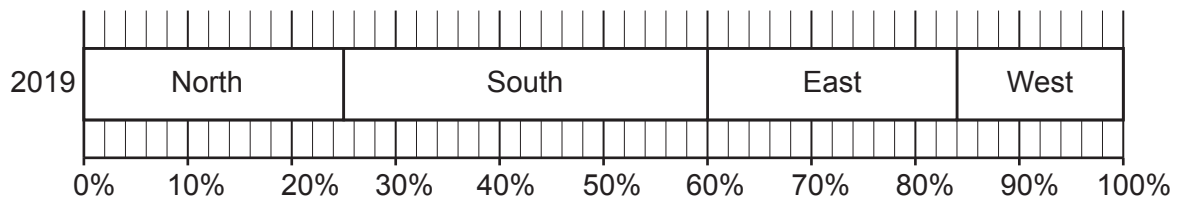
(i) car occupants,

..... [2]

(ii) motorcyclists and cyclists combined.

..... [2]

In country D, road deaths were also categorised by the part of the country in which the deaths occurred. This information is shown in the percentage bar chart.



(c) Find the number of road deaths which occurred in the East.

..... [2]

(d) Estimate the number of road deaths suffered by pedestrians in the South, stating also any assumption you must make in obtaining your answer.

Number

Assumption

..... [4]

[Question 10 continues on the next page]

To make fair comparisons between different places, standardised rates are calculated using a standard population.

(e) For death rates, the standard population usually consists of data showing the percentages of the population in different age group categories.

(i) Explain why such a standard population is used.

.....
..... [1]

(ii) Explain why, for standardised **road accident** death rates, such a standard population might **not** be appropriate.

.....
.....
.....
..... [2]

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced online in the Cambridge Assessment International Education Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download at www.cambridgeinternational.org after the live examination series.

Cambridge Assessment International Education is part of Cambridge Assessment. Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate (UCLES), which is a department of the University of Cambridge.