Cambridge Assessment International Education

Cambridge O Level

CANDIDATE NAME									
CENTRE NUMBER						CANDIDATE NUMBER				

0205050175

STATISTICS **4040/13**

Paper 1 **October/November 2020**

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 **20** pages. Blank pages are indicated.

1 In a town there are twelve schools, of which eight are public and four are private.
A schools inspector is to select three of these schools for inspection.
She considers three different methods of sampling: random, systematic, and stratified (by school type).
The following sampling frame is used.

School	A	B	C	D	E	F	G	H	I	J	K	L
Type	public	public	public	public	public	public	public	public	private	private	private	private
Sampling number	01	02	03	04	05	06	07	08	09	10	11	12

(a) State which **one or more** of the considered methods could possibly have been used if she actually selected

(i) schools B, F and J,

..... [2]

(ii) schools D, E and I.

..... [1]

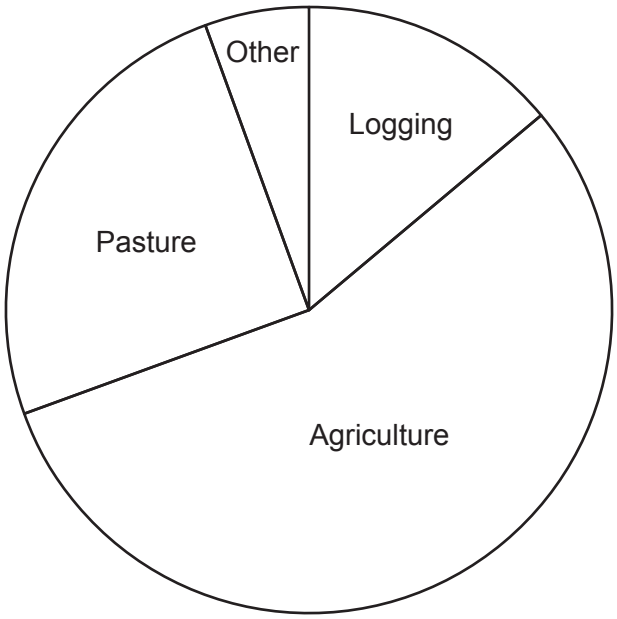
(b) State which **one** of the considered methods must certainly have been used if she actually selected schools A, B and C.

..... [1]

(c) If, instead of selecting three schools, the inspector carried out a census of schools in the town, state the number of schools she would select.

..... [1]

2 Over a period of time a country lost an area of 2.7 million hectares of its rainforest. The reasons for this loss were categorised as Logging, Agriculture, Pasture and Other. The pie chart, which is drawn to scale, represents the area of lost rainforest for each category.



Find

(a) the percentage of lost rainforest that was lost to Pasture,

..... [2]

(b) the area of rainforest that was lost to Agriculture.

..... million hectares [2]

3 At a research station in Antarctica, the external temperature is measured each day at noon. For one particular week the following values, in °C, were obtained.

–12 –14 –10 –8 –11 –16 –13

(a) For these values, find

(i) the mean,

..... [2]

(ii) the variance.

..... [2]

The internal temperature at the research station is kept constant at 20 °C.

(b) Write down the mean and range of the differences between the internal and external temperatures at noon for this week.

Mean = ...

Range = ... [2]

4 In the country of Laarfland the unit of currency is the 'chukkel'. The government suspects that 2% of the one-chukkel coins in circulation are counterfeit (not genuine), and the remainder are genuine.
Assume that the government's suspicion is accurate.

At a bakery, a shopper receives two of these coins in his change.

Find the probability that

(a) both coins are genuine,

..... [2]

(b) one coin is genuine and one coin is counterfeit.

..... [2]

At a pharmacy, another shopper receives three of these coins in her change.

(c) Find the probability that at least one coin is counterfeit.

..... [2]

5 Dan, Eva and Flo are movie fans. They discuss the preferences of their friends for watching three different types of movie: Adventure, Romance and Comedy.

Dan claims that girls prefer Romance over Adventure.
Eva claims that boys prefer Adventure over Comedy.
Flo claims that the popularity of Comedy is the same for girls and boys.

To investigate these claims they ask some of their friends about their preferences.

They record whether the friend asked is a girl (G) or boy (B), and whether they prefer Adventure (A), Romance (R), or Comedy (C).
Their raw data is as follows:

BC	GR	GC	BA	BR	GR		BC	GC	GR	BA	GC	GC
GA	BC	GA	BC	GC	GR		GC	BA	GR	GC	BC	BR

For example, the first friend asked was a boy who preferred Comedy.

(a) Summarise the data in a two-way table.

[4]

(b) Does the data support the claims of each of Dan, Eva and Flo? Justify your answers.

Dan .....

.....

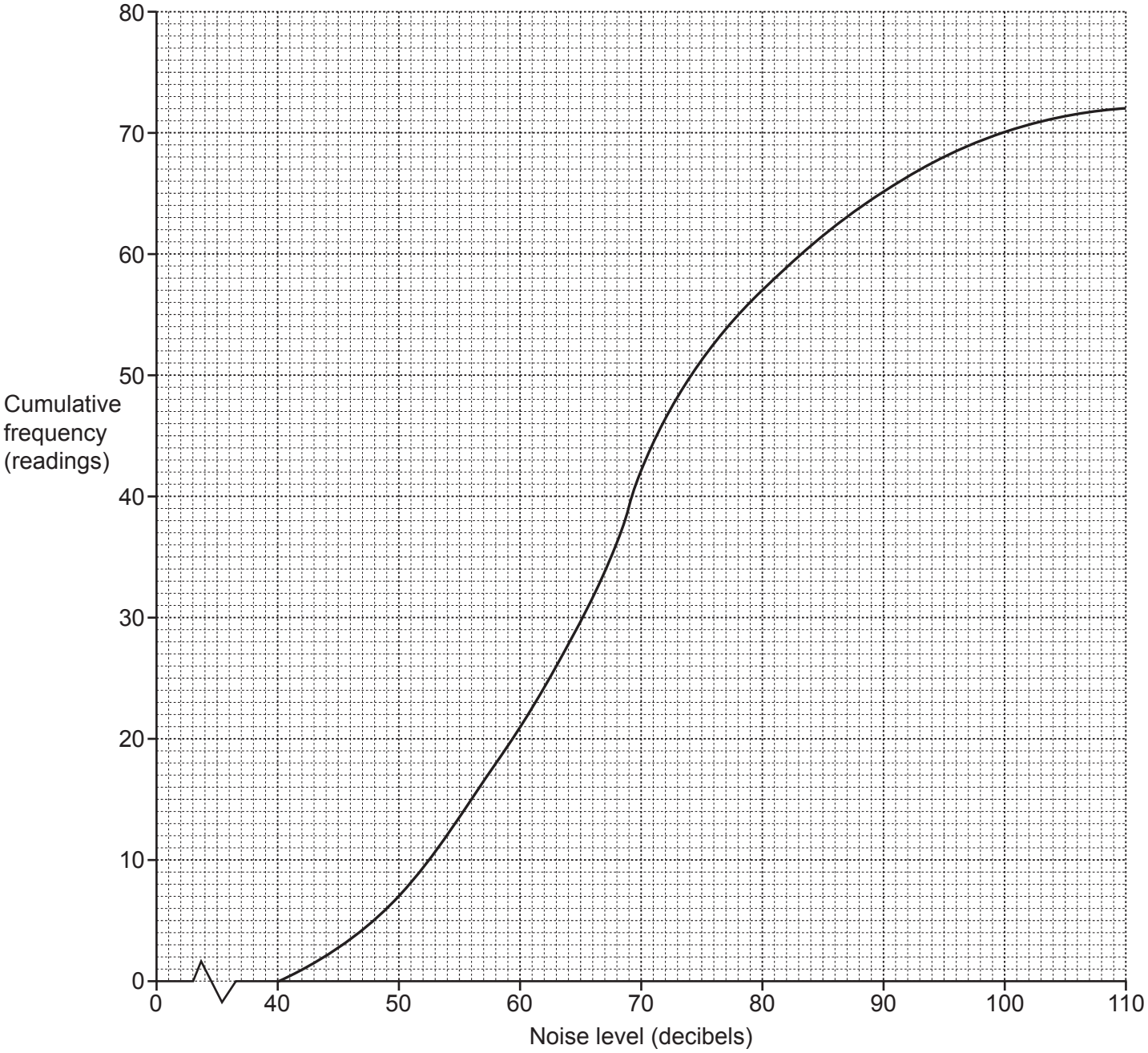
Eva .....

.....

Flo .....

..... [3]

6 Readings of the noise levels at a particular location in a city were recorded at hourly intervals over a period of three days. The 72 readings, in decibels, are illustrated in the cumulative frequency curve below.



(a) Use the graph to estimate, for these noise levels,

(i) the median,

..... decibels [1]

(ii) the interquartile range, given that the upper quartile is 77 decibels,

..... decibels [3]

(iii) the 36th percentile.

..... decibels [2]

An action group on noise pollution argues that there should be a law prohibiting noise levels in a city from exceeding 90 decibels for more than 7.5% of the time.

(b) Based on these readings, determine whether or not such a law would be broken in this city at present.

[2]

7 In this question, where required, calculate death rates per thousand of the population.

At the start of 2019 the population of a town was 17 500. During the year there were 147 deaths.

(a) Calculate the crude death rate for the town in 2019.

..... [2]

The table below gives further information on the town's population and deaths in 2019, together with the standard population of the region in which the town is situated.

Age group	*Percentage of population in age group*	*Age group death rate, per thousand*	*Standard population* (%)
0–19	16	2.5	20
20–39	22	6.0	25
40–59	30	7.2	30
60–79	24	10.5	20
Over 79	8	25.0	5

(b) Calculate the number of deaths in the 60–79 age group.

..... [3]

(c) Explain the purpose of calculating a standardised death rate.

.....

..... [1]

(d) Calculate the standardised death rate for the town in 2019.

..... [4]

(e) Explain why the standardised death rate is lower than the crude death rate in this case.

.....

..... [1]

The table below gives information on four other towns in the same region for 2019.

Town	*Population*	*Crude death rate, per thousand*	*Standardised death rate, per thousand*
P	15 700	10.4	8.9
Q	13 100	6.8	9.4
R	19 500	6.7	8.2
S	12 400	12.2	10.2

(f) By calculating the number of deaths in each of these four towns, find the town in which there were fewest deaths.

..... [3]

(g) Which of the four towns P, Q, R and S appears to have the highest proportion of young people? Explain your answer.

.....

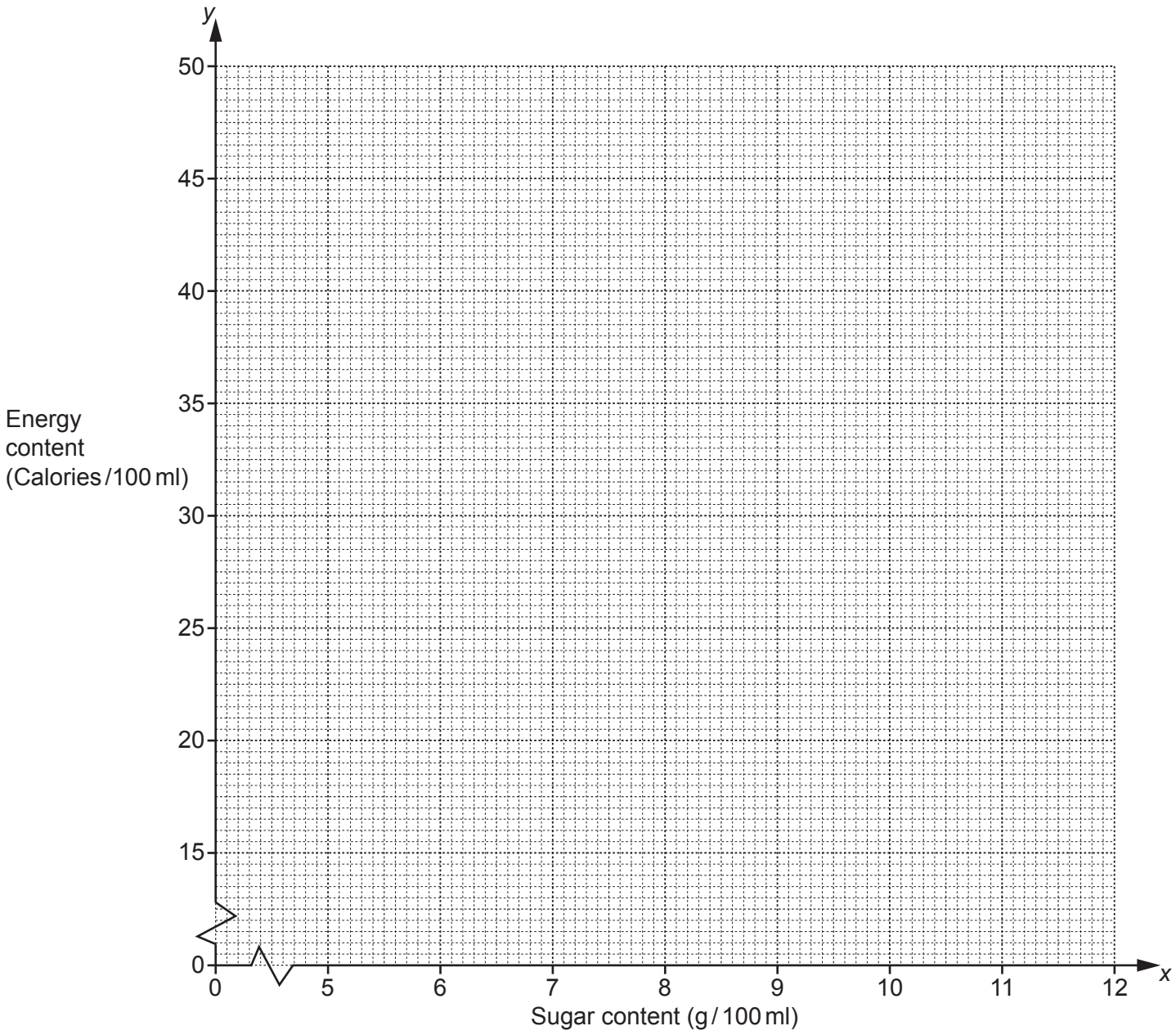
.....

..... [2]

8 Canned fizzy drinks often contain a lot of sugar, which is one source of energy in the human diet. For eight such drinks a scientist measured the sugar content, in grams per 100 millilitres of drink (g / 100 ml), and the energy content, in Calories per 100 millilitres of drink (Calories / 100 ml). Results are shown in the table below.

Drink	A	B	C	D	E	F	G	H
Sugar content, *x* (g / 100ml)	7.2	9.1	6.0	11.4	5.0	10.6	9.7	8.2
Energy content, *y* (Calories / 100ml)	29	39	27	47	24	44	40	34

(a) Plot these data on the grid below.



[2]

(b) Describe fully the correlation between the plotted points.

..... [2]

The data have an overall mean of (8.4, 35.5) and a lower semi-average of (6.6, 28.5).

(c) Find the upper 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) Use the line you have drawn in part **(d)** to estimate the energy content of 200 ml of a fizzy drink which contains 15 g of sugar.

.....Calories [2]

A nutritionist recommends that, for an adult, the maximum daily energy intake from sugar should be 125 Calories.

Assume that sugar is the only source of energy in fizzy drinks.

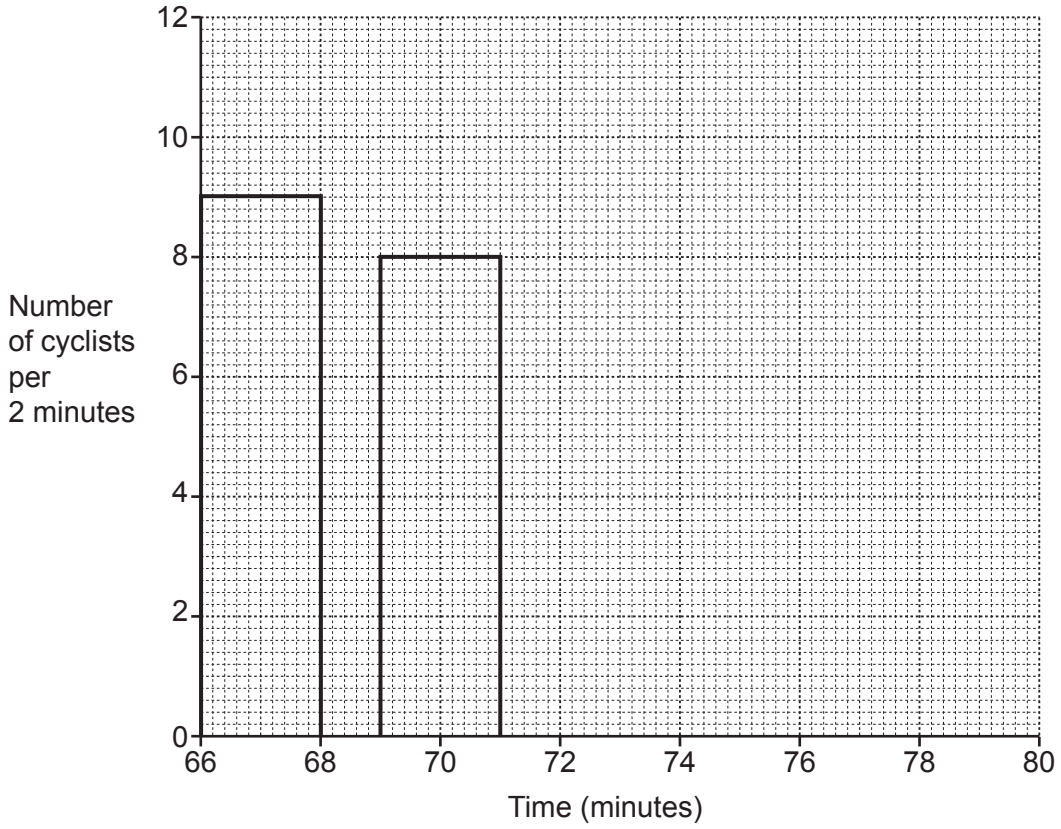
(f) For which, if any, of the drinks A–H would the consumption by an adult of one 330 ml can per day alone exceed this maximum? Justify your answer.

..... [3]

9 In a time-trial cycle race competitors ride around a course one at a time, and the times they take to complete the course are recorded. The following table summarises the results.

Time (minutes)	*Number of cyclists*
66–under 68	9
68–under 69	5
69–under 71	8
71–under 74	6
74–under 80	4

(a) On the grid below, draw a histogram to illustrate the data in this table. The rectangles representing the 66–under 68 class and 69–under 71 class have already been drawn for you.



[4]

A journalist, reporting on the race, described the cyclists with times between 69 minutes 30 seconds and 73 minutes as 'moderate sprinters'.

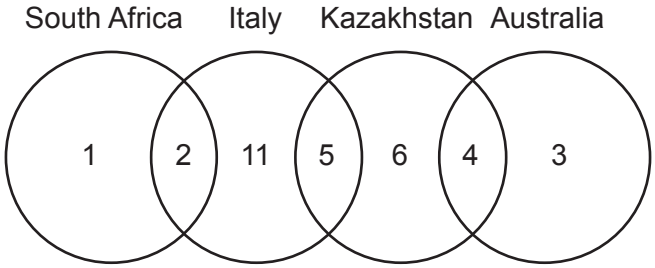
(b) Use your histogram to estimate the number, according to the journalist, of 'moderate sprinters' in the race. State any assumption made in obtaining your answer.

Number .....

Assumption .....

..... [3]

A cycling organisation, wishing to promote the sport world-wide, investigates the countries in which these cyclists have raced.
The diagram below shows the number of cyclists who have raced in one or more of the countries South Africa, Italy, Kazakhstan and Australia.



(c) Use this information to find the number of cyclists who have raced in

(i) South Africa or Kazakhstan,

..... [1]

(ii) South Africa and Kazakhstan,

..... [1]

(iii) one or more of Italy, Kazakhstan or Australia.

..... [1]

A cyclist is chosen at random to make a promotional video.

(d) Find the probability of choosing a cyclist who has raced in

(i) South Africa and Italy or Kazakhstan and Australia,

..... [1]

(ii) Italy, given that the cyclist has raced in South Africa,

..... [1]

(iii) South Africa, given that the cyclist has raced in Italy.

..... [1]

Two cyclists are chosen at random to be interviewed on their country preferences for racing.

(e) Find the probability of choosing cyclists who, by combining their experiences, have raced in all four of these countries.

..... [3]

10 Lerato and Modise use the social networking service 'Disputandum'. The service allows users to send messages with lengths of up to 90 characters (letters, spaces, punctuation marks, etc.).

Lerato recorded the lengths of the messages she sent during one particular week and summarised her results in the following table.

Message length (number of characters)	*Number of messages*	
1–15	2	
16–30	3	
31–45	10	
46–60	11	
61–75	9	
76–90	3	

(a) State the modal class.

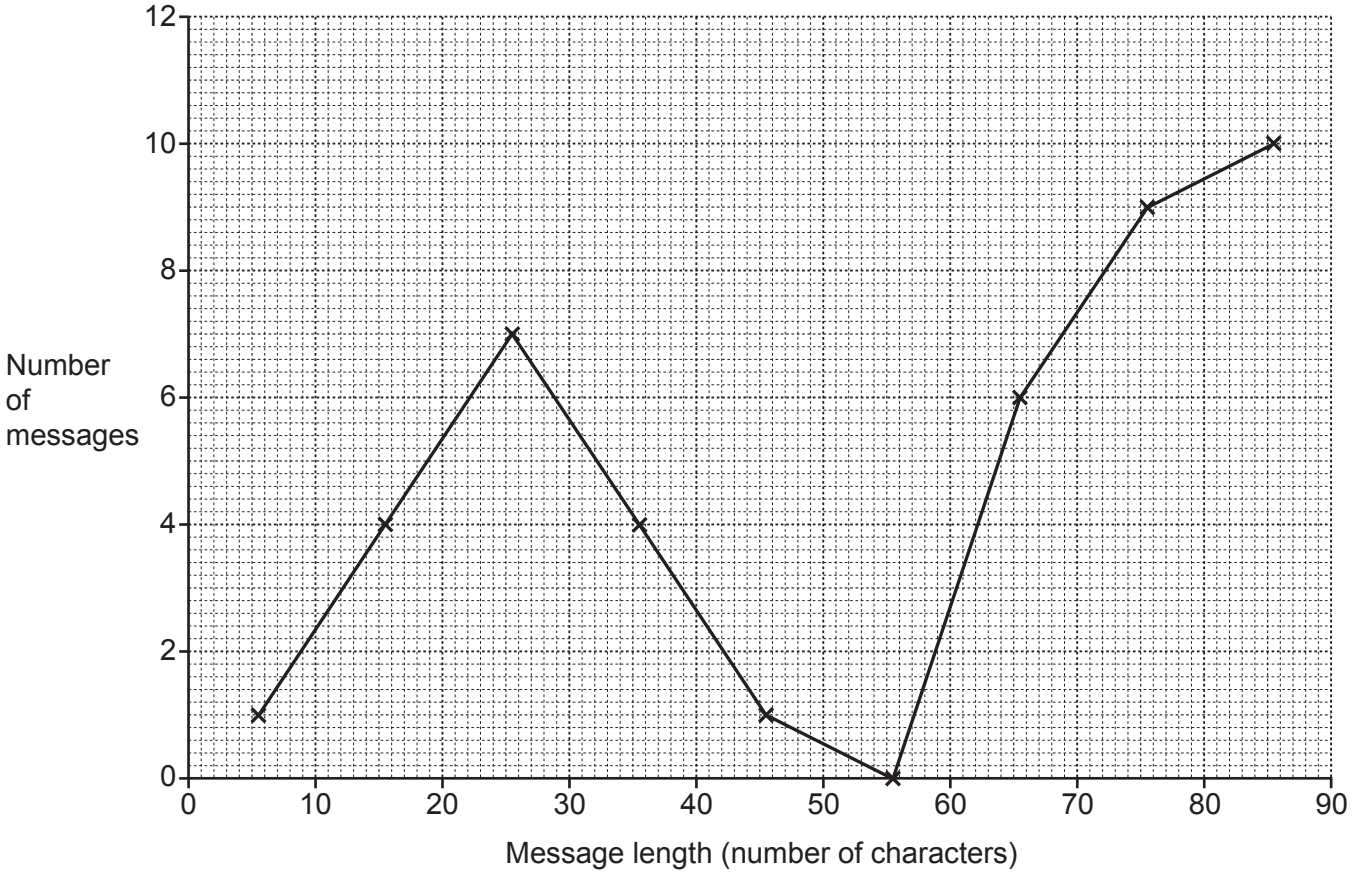
..... [1]

(b) Estimate the mean and standard deviation of these message lengths.
Give your answers correct to 3 significant figures.

Mean = ... characters

Standard deviation = ... characters [7]

Modise also recorded the lengths of the messages he sent during the same week. After grouping his results using a constant but different class interval from that used by Lerato, he summarised them in the following frequency polygon.



(c) State the class interval which Modise used when grouping his results.

..... [1]

(d) Find the difference between the total number of messages sent by Lerato and Modise during this week, stating also which of the two sent the greater number.

..... [2]

(e) Draw on the grid above a frequency polygon to illustrate Lerato's results. [3]

(f) Make two comparisons between the lengths of the messages sent by Lerato and Modise during this week.

1 .....

.....

2 .....

..... [2]

BLANK PAGE

BLANK PAGE