Cambridge Assessment International Education

Cambridge O Level

CANDIDATE NAME									
CENTRE NUMBER						CANDIDATE NUMBER				

3826016749

STATISTICS **4040/12**

Paper 1 **October/November 2023**

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 each of the following, state the name of a method of sampling in which

(a) a sampling frame is not required,

..... [1]

(b) the individual items are selected at regular intervals from a sampling frame,

..... [1]

(c) the sample is selected so that the proportions of different categories in the sample correspond with those of the population,

..... [1]

(d) the sampling method is free from bias.

..... [1]

2 The members of a reading group meet periodically to discuss the books they have read. At the start of one meeting the group leader questions the eleven attendees on the number of books they have read since the previous meeting. She obtains the following replies.

3 2 4 2 7 10 6 2 9 5 2

(a) Find

(i) the mode,

..... [1]

(ii) the median,

..... [2]

(iii) the upper quartile.

..... [1]

(b) Explain why the mode is not a good measure of central tendency in this case.

.....

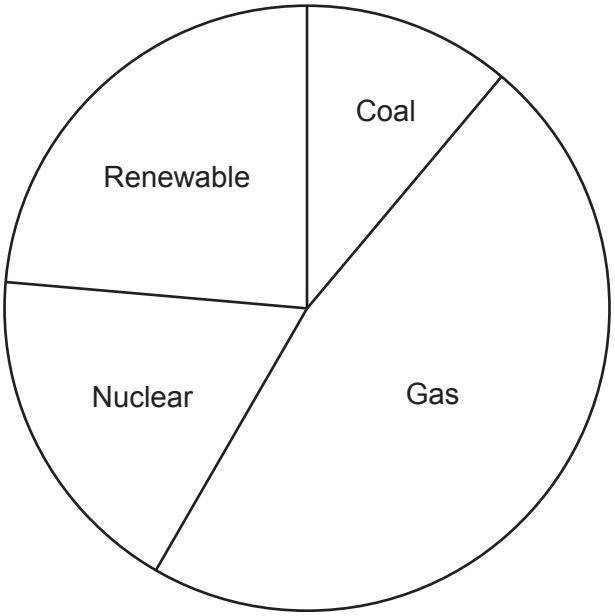
..... [1]

Two members of the group were late for the meeting and missed the group leader's question at the start. Their replies would have been 3 books and 8 books. If this extra data is now included in the data above, and the measures in part **(a)** are found for all thirteen attendees, the value of only one measure would change.

(c) State the measure and its changed value.

..... [1]

3 In a particular country there are two electricity supply companies, Energen and Powerlec. They each generate electricity from four sources: Coal, Gas, Nuclear, and Renewable. For Energen the proportions from each of these is shown in the pie chart, of radius 4 cm, which is drawn to scale.



Use the chart to find, for Energen, the percentage of electricity

(a) generated from Coal,

..... [2]

(b) **not** generated from Gas.

..... [2]

The total amount of electricity generated by Powerlec is three times the total amount of electricity generated by Energen.

(c) If a comparative pie chart is drawn for Powerlec, find, correct to 1 decimal place, its radius.

..... [3]

4 Agnetha, a Statistics student in Sweden, is interested in weather patterns in her town. Over a period of 20 days she records the weather each day in the morning, afternoon and evening, as one of mainly sun (S), cloud (C) or rain (R). Her raw data is as follows.

CSS	CSS	SCS	SSS	RRR	RRR	RRR	CSC	CSS	SCS
SSS	CSS	RSS	SSS	CSR	CRS	CCS	SSS	CSS	CRC

For example, on the first and second days, there was mainly cloud in the morning, sun in the afternoon, and sun in the evening.

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

[4]

(b) Use your table to write down **two** observations about weather patterns in this period.

1 .....

.....

2 .....

.....

[2]

Agnetha's classmate Bjorn says that a disadvantage of her doing a two-way table is that the original data is lost.

(c) Write down **one** observation about weather patterns in this period that can be made from the raw data which cannot be made from the table.

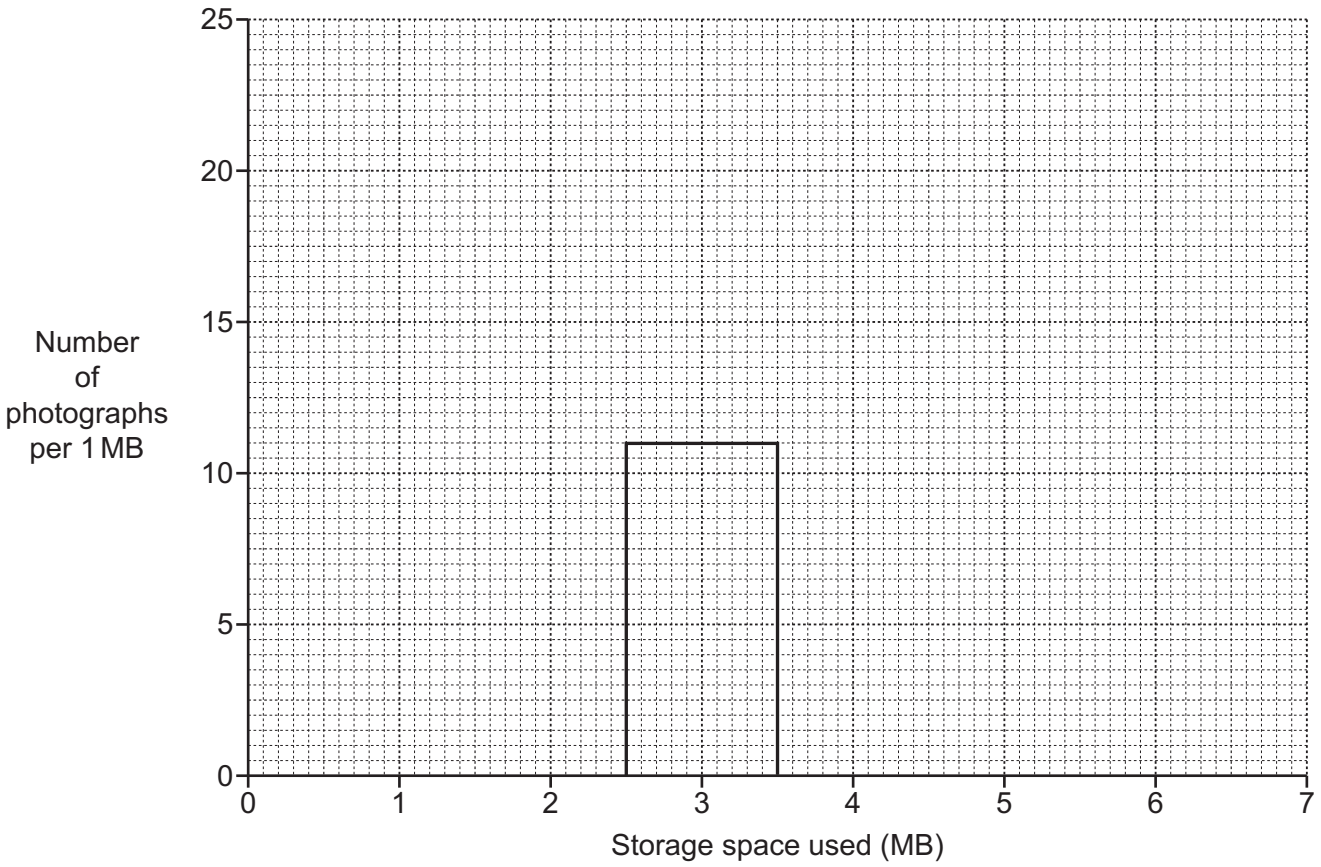
.....

..... [1]

5 Andrew takes many photographs on his new mobile phone. The amount of storage space used on the phone, in megabytes (MB), by the last 50 photographs taken, is summarised in the table.

Storage space used (MB)	*Number of photographs*
1.0–under 2.5	6
2.5–under 3.5	11
3.5–under 4.0	9
4.0–under 5.0	14
5.0–under 7.0	10

(a) On the grid below, draw a histogram to illustrate the data in the table. The rectangle representing the 2.5–under 3.5 class has already been drawn for you.



[5]

(b) Use the table to estimate, in megabytes, the total amount of storage space used on Andrew's phone by these 50 photographs.

..... MB [3]

6 Tony and Cleo are the two hairdressers who work in a salon. All customers at the salon must book their appointments beforehand. From experience it is known that 10% of customers fail to appear for their appointments.

(a) One morning Tony has 3 booked appointments and Cleo has 2 booked appointments.

Find the probability that, on this morning,

(i) only Cleo's second customer fails to appear,

..... [2]

(ii) both of Cleo's customers appear, but one of Tony's does not.

..... [3]

All customers are offered a drink when they arrive at the salon. From experience it is known that 75% of customers accept the drink.

(b) One afternoon Tony has 2 booked appointments and Cleo has 1 booked appointment.

Find the probability that, on this afternoon, 3 drinks are served to customers.

..... [2]

7 An animal sanctuary cares for young animals until they are older and can safely be returned to the wild.
The table shows the recorded height of a baby giraffe, given the name Zara, over the first 400 days of her life.

Age, x (days)	50	100	150	200	250	300	350	400
Height, y (cm)	188	206	225	237	254	266	273	279

The data have an overall mean of (225, 241), and a lower semi-average of (125, 214).

(a) Find the upper semi-average of the data.

..... [2]

(b) Find the equation of the line of best fit to the data in the form $y = mx + c$.

..... [3]

(c) Use your equation to estimate, to the nearest centimetre, Zara's height

(i) at birth,

..... [1]

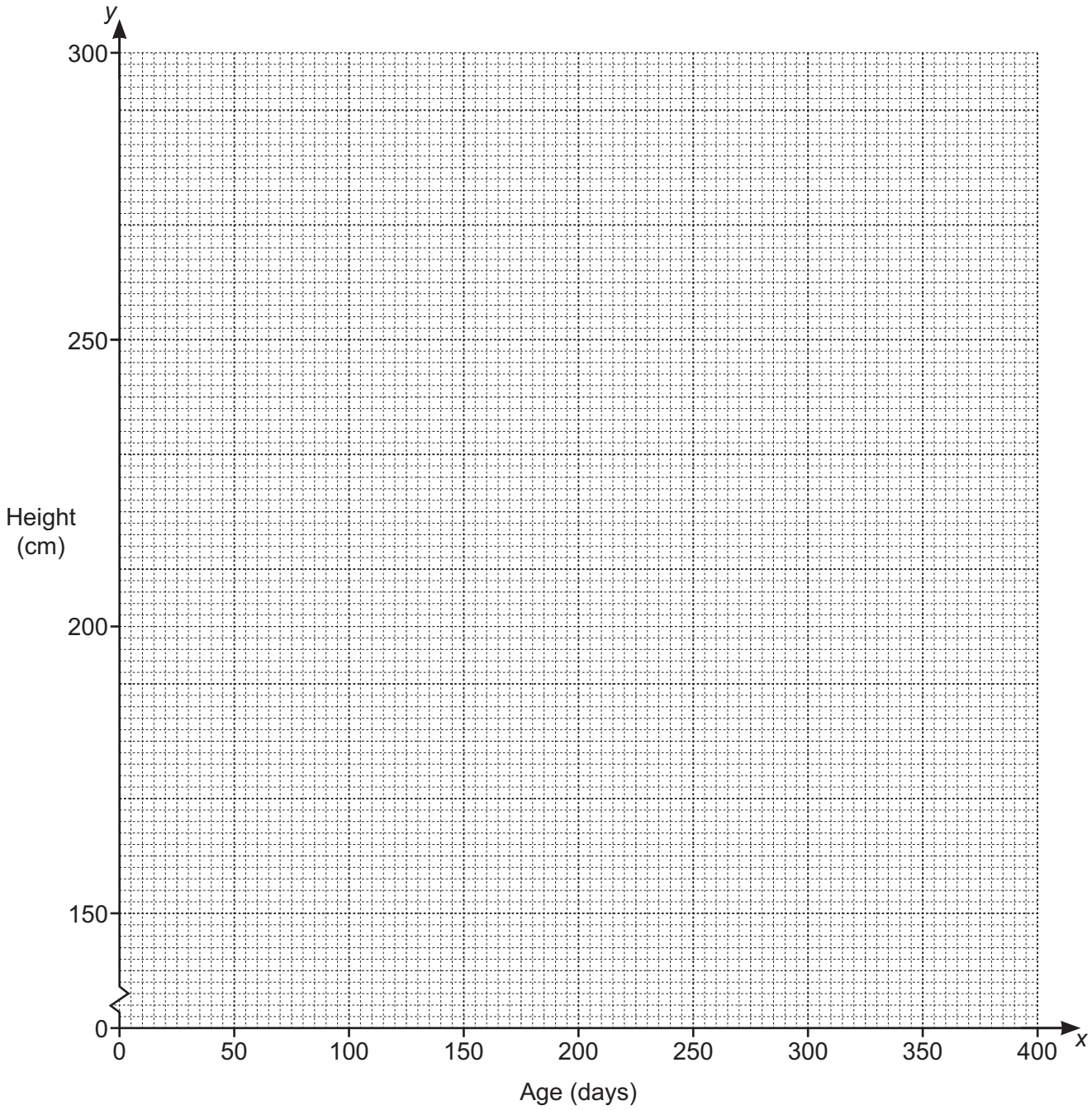
(ii) at age 600 days.

..... [2]

(d) On the grid opposite, plot the data given in the table above. [2]

(e) Draw on the grid the line whose equation you found in part **(b)** for ages between 0 and 400 days.

[2]



(f) By inspecting the plotted points, explain why it might be considered that it was inappropriate to find a line of best fit in the form $y = mx + c$ in this case.

.....

..... [1]

(g) State how you would expect Zara's actual height at 600 days to compare with the value calculated in part **(c)(ii)**.

..... [1]

8 In this question the goal-scoring rate for a footballer is defined as the number of goals scored per 1000 minutes played.

Pip Gladiola is the manager of Statstown Rovers football team. He analyses the performance of members of his squad by calculating their goal-scoring rates.
Over a particular period of time his team captain, Razak, has played 1875 minutes and scored 8 goals.

(a) Show that Razak's crude goal-scoring rate, correct to 1 decimal place, is 4.3.

[1]

Pip knows that a factor affecting the goal-scoring performance of a player is the quality of the opposition he faces. He standardises goal-scoring rates to allow for this.
The table shows information for Razak over this period of time, together with the standard population for all matches played.

Opposition quality	*Number of goals*	*Minutes played*	*Opposition quality goal-scoring rate*	*Standard population of opposition quality* (%)
High	1	625		30
Moderate	3	750		50
Weak	4	500		20

(b) Calculate Razak's goal-scoring rate for each opposition quality group and insert the values into the table.

[3]

(c) Use your results from part **(b)** to calculate Razak's standardised goal-scoring rate.

..... [3]

Pip has four players in his squad, called 'strikers', whose special responsibility is to score goals. He never selects all of them for the start of a match, but may remove and substitute a player during a match.
The table gives information on the strikers' goal-scoring rates, correct to 1 decimal place.

Striker	*Minutes played*	*Standardised goal-scoring rate*	*Crude goal-scoring rate*
Alonso	1540	10.1	8.4
Benjani	1370	7.9	10.2
Camara	1680	6.6	6.5
Diame	1050	9.3	11.4

(d) Identify the striker and the number of goals scored by

(i) the striker who scored most goals,

..... [2]

(ii) the striker who scored fewest goals.

..... [1]

(e) Identify, with a reason, the striker who Pip would judge to have performed best over this period of time.

.....

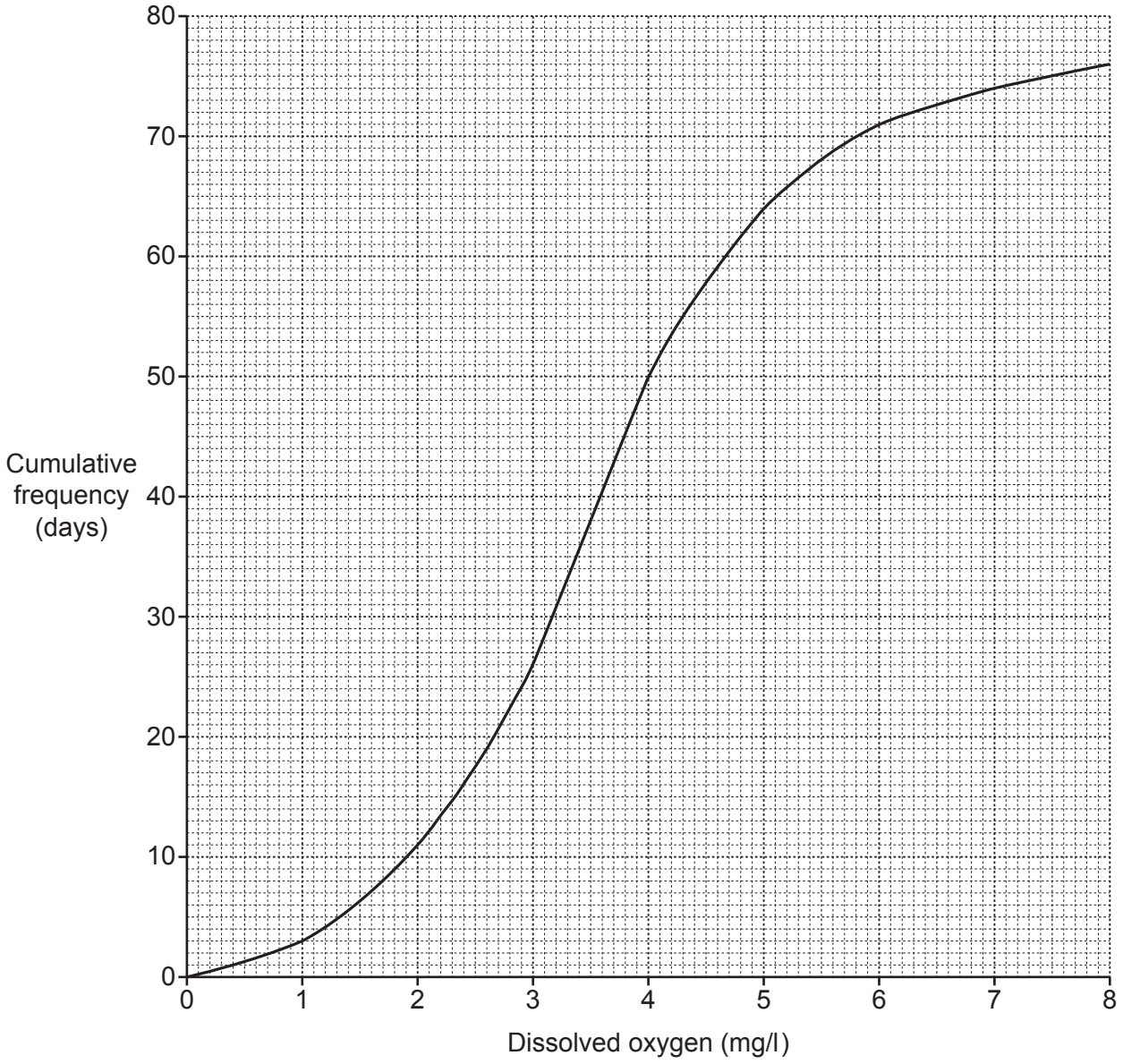
..... [2]

For a charity match Pip selects, at random, from his strikers, two to play in the first half of the match. He removes them at half time and substitutes the other two strikers to play in the second half.

(f) Find the probability that Alonso and Diame play in the first half of the match, and Benjani and Camara play in the second half.

..... [3]

9 The suitability of water in a river for supporting aquatic life (fish etc.) is assessed by the quantity of dissolved oxygen in it. For a particular river, Hadiya, an environmental scientist, measured this quantity in milligrams per litre (mg/l) over a period of 76 days. Each day she took one measurement from the same place in the river.
Her results are summarised in the cumulative frequency curve below.



(a) Use the graph to estimate

(i) the median,

..... mg/l [1]

(ii) the interquartile range, given that the lower quartile is 2.6 mg/l,

..... mg/l [3]

(iii) the 35th percentile.

..... mg/l [2]

Hadiya categorises the suitability of river water for supporting aquatic life as follows.

Dissolved oxygen (mg/l)	under 4.0	4.0–under 6.0	6.0–under 8.0
Suitability	unsuitable	marginal	adequate

(b) Use the graph to estimate

(i) the percentage of the time Hadiya considered the water to have been 'adequate',

..... [2]

(ii) the median dissolved oxygen value during the times Hadiya considered the water to have been 'marginal'.

..... mg/l [2]

Hadiya believes that if river water is categorised as 'unsuitable' for more than 30% of the time, then action must be taken to improve it.

(c) **(i)** Show that for this river at present Hadiya will believe such action to be necessary.

[2]

(ii) If such action were taken, and it resulted in all dissolved oxygen values being increased by 1.5 mg/l, determine whether or not Hadiya would believe even further action should be taken.

..... [2]

(d) Explain how Hadiya might have improved on her sampling of the river water.

.....

.....

.....

..... [2]

10 A university offers short online courses to members of the general public. The length of a course is indicated by the hours of study needed to complete it.
The table shows the lengths of the 45 such courses offered by the Science department.

Course length, x (hours)	*Number of courses, f*	
2	4	
5	6	
8	7	
12	13	
14	8	
16	4	
18	3	

For these course lengths,

(a) find the range,

..... [1]

(b) calculate the mean and standard deviation, giving your answers to 3 significant figures.

Mean .....

Standard deviation .....
[6]

Short online courses are also offered by four other departments. The table shows measures of course length in these departments.

Department	*Mean* (hours)	*Median* (hours)	*Standard deviation* (hours)
Business	12.4	12	4.67
Education	12.8	12	2.94
History	12.0	11	2.85
Languages	12.5	10	3.74

(c) State, giving the reason for your choice, in which of the four departments Business, Education, History and Languages,

(i) courses are generally longest,

..... [1]

(ii) courses are generally most similar in length,

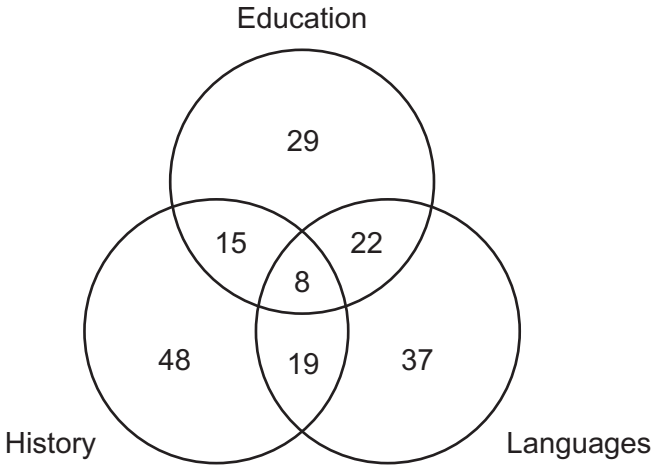
..... [1]

(iii) the median may be a more appropriate measure of central tendency than the mean.

..... [1]

[Question 10 continues on the next page]

The courses offered in the Education, History and Languages departments are all free of charge. The diagram shows the number of people who registered for courses in one or more of these departments in one particular month.



(d) Find the number of people who registered in

(i) Languages,

..... [1]

(ii) Education or History or both,

..... [1]

(iii) exactly two of these departments.

..... [1]

(e) Calculate the mean number of these departments in which these people registered during this month.

..... [3]