Answer **all** questions.

1 At a certain small restaurant, the waiting time is defined as the time between sitting down at a table and a waiter first arriving at the table. This waiting time is dependent upon the number of other customers already seated in the restaurant.

Alex is a customer who visited the restaurant on 10 separate days. The table shows, for each of these days, the number, x, of customers already seated and his waiting time, y minutes.

x	9	3	4	10	8	12	7	11	2	6
y	11	6	5	11	9	13	9	12	4	7

(a) Calculate the equation of the least squares regression line of y on x in the form $y = a + bx$. *(4 marks)*

(b) Give an interpretation, in context, for each of your values of a and b. *(2 marks)*

(c) Use your regression equation to estimate Alex's waiting time when the number of customers already seated in the restaurant is:

(i) 5;

(ii) 25. *(2 marks)*

(d) Comment on the likely reliability of **each** of your estimates in part (c), given that, for the regression line calculated in part (a), the values of the 10 residuals lie between $+1.1$ minutes and -1.1 minutes. *(3 marks)*

2 Xavier, Yuri and Zara attend a sports centre for their judo club's practice sessions. The probabilities of them arriving late are, independently, 0.3, 0.4 and 0.2 respectively.

(a) Calculate the probability that for a particular practice session:

(i) all three arrive late; *(1 mark)*

(ii) none of the three arrives late; *(2 marks)*

(iii) exactly one of the three arrives late. *(3 marks)*

(b) Zara's friend, Wei, also attends the club's practice sessions. The probability that Wei arrives late is 0.9 when Zara arrives late, and is 0.25 when Zara does not arrive late.

Calculate the probability that for a particular practice session:

(i) both Zara and Wei arrive late; *(2 marks)*

(ii) either Zara or Wei, but not both, arrives late. *(3 marks)*

3 When an alarm is raised at a market town's fire station, the fire engine cannot leave until at least five fire-fighters arrive at the station. The call-out time, X minutes, is the time between an alarm being raised and the fire engine leaving the station.

The value of X was recorded on a random sample of 50 occasions. The results are summarised below, where $\bar{x}$ denotes the sample mean.

$$\sum x = 286.5 \qquad \sum (x - \bar{x})^2 = 45.16$$

(a) Find values for the mean and standard deviation of this sample of 50 call-out times. *(2 marks)*

(b) Hence construct a 99% confidence interval for the mean call-out time. *(4 marks)*

(c) The fire and rescue service claims that the station's mean call-out time is less than 5 minutes, whereas a parish councillor suggests that it is more than $6\frac{1}{2}$ minutes.

Comment on **each** of these claims. *(2 marks)*

4 [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows the times, in seconds, taken by a random sample of 10 boys from a junior swimming club to swim 50 metres freestyle and 50 metres backstroke.

Boy	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Freestyle (x seconds)	30.2	32.8	25.1	31.8	31.2	35.6	32.4	38.0	36.1	34.1
Backstroke (y seconds)	33.5	35.4	37.4	27.2	34.7	38.2	37.7	41.4	42.3	38.4

(a) On **Figure 1**, complete the scatter diagram for these data. *(2 marks)*

(b) Hence:

(i) give **two** distinct comments on what your scatter diagram reveals; *(2 marks)*

(ii) state, **without calculation**, which of the following 3 values is most likely to be the value of the product moment correlation coefficient for the data in your scatter diagram.

0.912 0.088 0.462 *(1 mark)*

(c) In the sample of 10 boys, one boy is a junior-champion freestyle swimmer and one boy is a junior-champion backstroke swimmer.

Identify the **two** most likely boys. *(2 marks)*

(d) **Removing the data for the two boys whom you identified in part (c):**

(i) calculate the value of the product moment correlation coefficient for the remaining 8 pairs of values of x and y; *(3 marks)*

(ii) comment, in context, on the value that you obtain. *(1 mark)*

5 (a) The baggage loading time, X minutes, of a chartered aircraft at its UK airport may be modelled by a normal random variable with mean 55 and standard deviation 8.

Determine:

(i) $\mathrm{P}(X < 60)$; *(3 marks)*

(ii) $\mathrm{P}(55 < X < 60)$. *(2 marks)*

(b) The baggage loading time, Y minutes, of a chartered aircraft at its overseas airport may be modelled by a normal random variable with mean μ and standard deviation 16.

Given that $\mathrm{P}(Y < 90) = 0.95$, find the value of μ. *(4 marks)*

6 The table shows, for a particular population, the proportion of people in each of the four main blood groups.

Blood group	**O**	**A**	**B**	**AB**
Proportion	0.40	0.28	0.20	0.12

(a) A random sample of 20 people is selected from this population.

Determine the probability that the sample contains:

(i) at most 10 people with blood group O; *(2 marks)*

(ii) exactly 3 people with blood group A; *(3 marks)*

(iii) more than 4 but fewer than 8 people with blood group B. *(3 marks)*

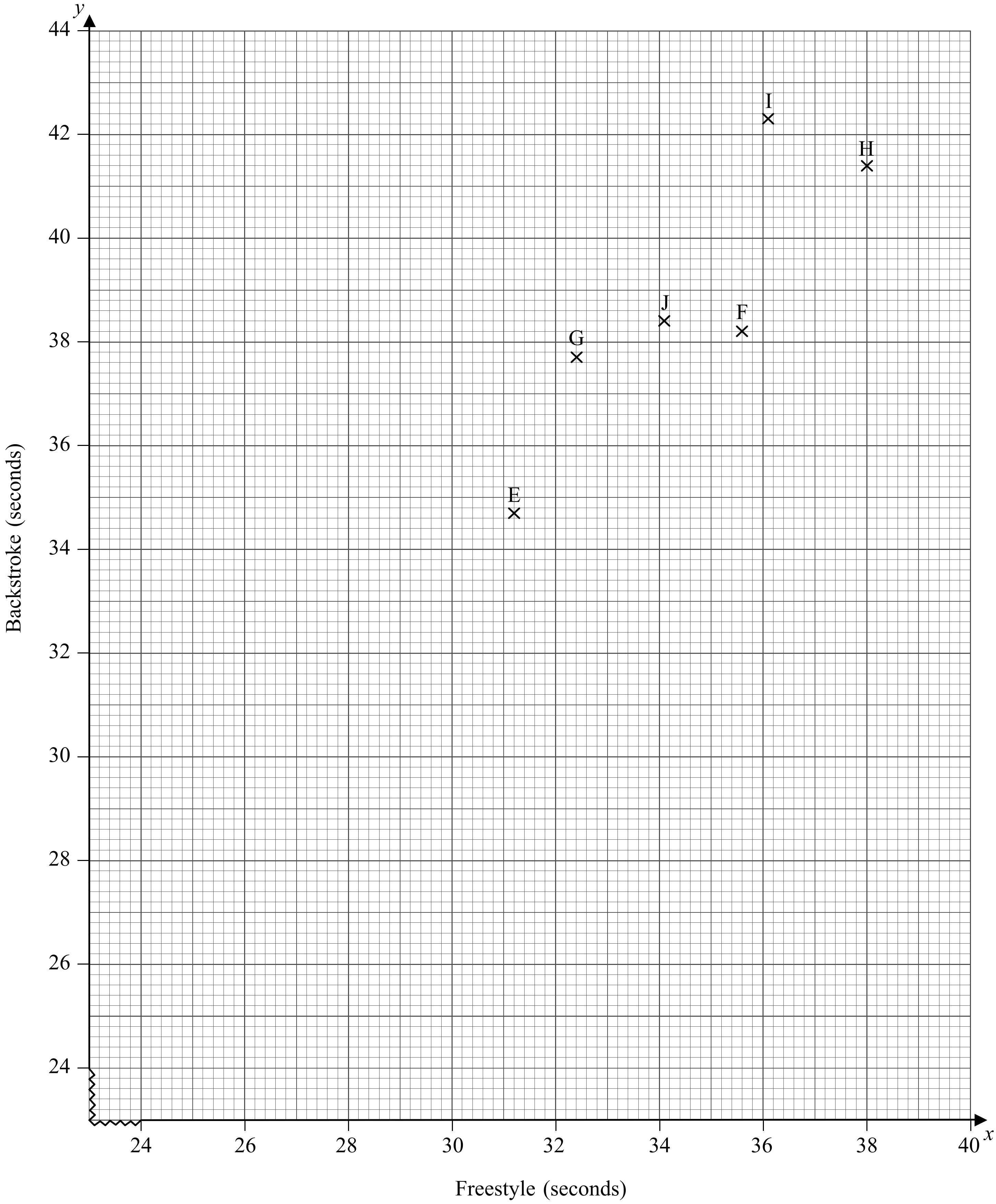
(b) A random sample of 500 people is selected from this population.

Find values for the mean and variance of the number of people in the sample with blood group AB. *(2 marks)*

END OF QUESTIONS

Figure 1 (for use in Question 4)

Scatter Diagram for Freestyle and Backstroke Swimming Times



Answer **all** questions.

1 At a certain small restaurant, the waiting time is defined as the time between sitting down at a table and a waiter first arriving at the table. This waiting time is dependent upon the number of other customers already seated in the restaurant.

Alex is a customer who visited the restaurant on 10 separate days. The table shows, for each of these days, the number, x, of customers already seated and his waiting time, y minutes.

x	9	3	4	10	8	12	7	11	2	6
y	11	6	5	11	9	13	9	12	4	7

(a) Calculate the equation of the least squares regression line of y on x in the form $y = a + bx$. *(4 marks)*

(b) Give an interpretation, in context, for each of your values of a and b. *(2 marks)*

(c) Use your regression equation to estimate Alex's waiting time when the number of customers already seated in the restaurant is:

(i) 5;

(ii) 25. *(2 marks)*

(d) Comment on the likely reliability of **each** of your estimates in part (c), given that, for the regression line calculated in part (a), the values of the 10 residuals lie between +1.1 minutes and −1.1 minutes. *(3 marks)*

2 Xavier, Yuri and Zara attend a sports centre for their judo club's practice sessions. The probabilities of them arriving late are, independently, 0.3, 0.4 and 0.2 respectively.

(a) Calculate the probability that for a particular practice session:

(i) all three arrive late; *(1 mark)*

(ii) none of the three arrives late; *(2 marks)*

(iii) only Zara arrives late. *(2 marks)*

(b) Zara's friend, Wei, also attends the club's practice sessions. The probability that Wei arrives late is 0.9 when Zara arrives late, and is 0.25 when Zara does not arrive late.

Calculate the probability that for a particular practice session:

(i) both Zara and Wei arrive late; *(2 marks)*

(ii) either Zara or Wei, but not both, arrives late. *(3 marks)*

3 When an alarm is raised at a market town's fire station, the fire engine cannot leave until at least five fire-fighters arrive at the station. The call-out time, X minutes, is the time between an alarm being raised and the fire engine leaving the station.

The value of X was recorded on a random sample of 50 occasions. The results are summarised below, where $\bar{x}$ denotes the sample mean.

$$\sum x = 286.5 \qquad \sum (x - \bar{x})^2 = 45.16$$

(a) Find values for the mean and standard deviation of this sample of 50 call-out times. *(2 marks)*

(b) Hence construct a 99% confidence interval for the mean call-out time. *(4 marks)*

(c) The fire and rescue service claims that the station's mean call-out time is less than 5 minutes, whereas a parish councillor suggests that it is more than $6\frac{1}{2}$ minutes.

Comment on **each** of these claims. *(2 marks)*

4 The time, x seconds, spent by each of a random sample of 100 customers at an automatic teller machine (ATM) is recorded. The times are summarised in the table.

Time (seconds)	**Number of customers**
$20 < x \leqslant 30$	2
$30 < x \leqslant 40$	7
$40 < x \leqslant 60$	18
$60 < x \leqslant 80$	27
$80 < x \leqslant 100$	23
$100 < x \leqslant 120$	13
$120 < x \leqslant 150$	7
$150 < x \leqslant 180$	3
Total	**100**

(a) Calculate estimates for the mean and standard deviation of the time spent at the ATM by a customer. *(4 marks)*

(b) The mean time spent at the ATM by a random sample of **36** customers is denoted by $\overline{Y}$.

(i) State why the distribution of $\overline{Y}$ is approximately normal. *(1 mark)*

(ii) Write down estimated values for the mean and standard error of $\overline{Y}$. *(2 marks)*

(iii) Hence estimate the probability that $\overline{Y}$ is less than $1\frac{1}{2}$ minutes. *(3 marks)*

5 [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows the times, in seconds, taken by a random sample of 10 boys from a junior swimming club to swim 50 metres freestyle and 50 metres backstroke.

Boy	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Freestyle (x seconds)	30.2	32.8	25.1	31.8	31.2	35.6	32.4	38.0	36.1	34.1
Backstroke (y seconds)	33.5	35.4	37.4	27.2	34.7	38.2	37.7	41.4	42.3	38.4

(a) On **Figure 1**, complete the scatter diagram for these data. *(2 marks)*

(b) Hence:

(i) give **two** distinct comments on what your scatter diagram reveals; *(2 marks)*

(ii) state, **without calculation**, which of the following 3 values is most likely to be the value of the product moment correlation coefficient for the data in your scatter diagram.

0.912 0.088 0.462 *(1 mark)*

(c) In the sample of 10 boys, one boy is a junior-champion freestyle swimmer and one boy is a junior-champion backstroke swimmer.

Identify the **two** most likely boys. *(2 marks)*

(d) **Removing the data for the two boys whom you identified in part (c):**

(i) calculate the value of the product moment correlation coefficient for the remaining 8 pairs of values of x and y; *(3 marks)*

(ii) comment, in context, on the value that you obtain. *(1 mark)*

6 Plastic clothes pegs are made in various colours.

The number of red pegs may be modelled by a binomial distribution with parameter p equal to 0.2.

The contents of packets of 50 pegs of mixed colours may be considered to be random samples.

(a) Determine the probability that a packet contains:

(i) less than or equal to 15 red pegs; *(2 marks)*

(ii) exactly 10 red pegs; *(2 marks)*

(iii) more than 5 but fewer than 15 red pegs. *(3 marks)*

(b) Sly, a student, claims to have counted the number of red pegs in each of 100 packets of 50 pegs. From his results the following values are calculated.

Mean number of red pegs per packet = 10.5

Variance of number of red pegs per packet = 20.41

Comment on the validity of Sly's claim. *(4 marks)*

7 (a) The weight, X grams, of soup in a carton may be modelled by a normal random variable with mean 406 and standard deviation 4.2.

Find the probability that the weight of soup in a carton:

(i) is less than 400 grams; *(3 marks)*

(ii) is between 402.5 grams and 407.5 grams. *(4 marks)*

(b) The weight, Y grams, of chopped tomatoes in a tin is a normal random variable with mean μ and standard deviation σ.

(i) Given that $P(Y < 310) = 0.975$, explain why:

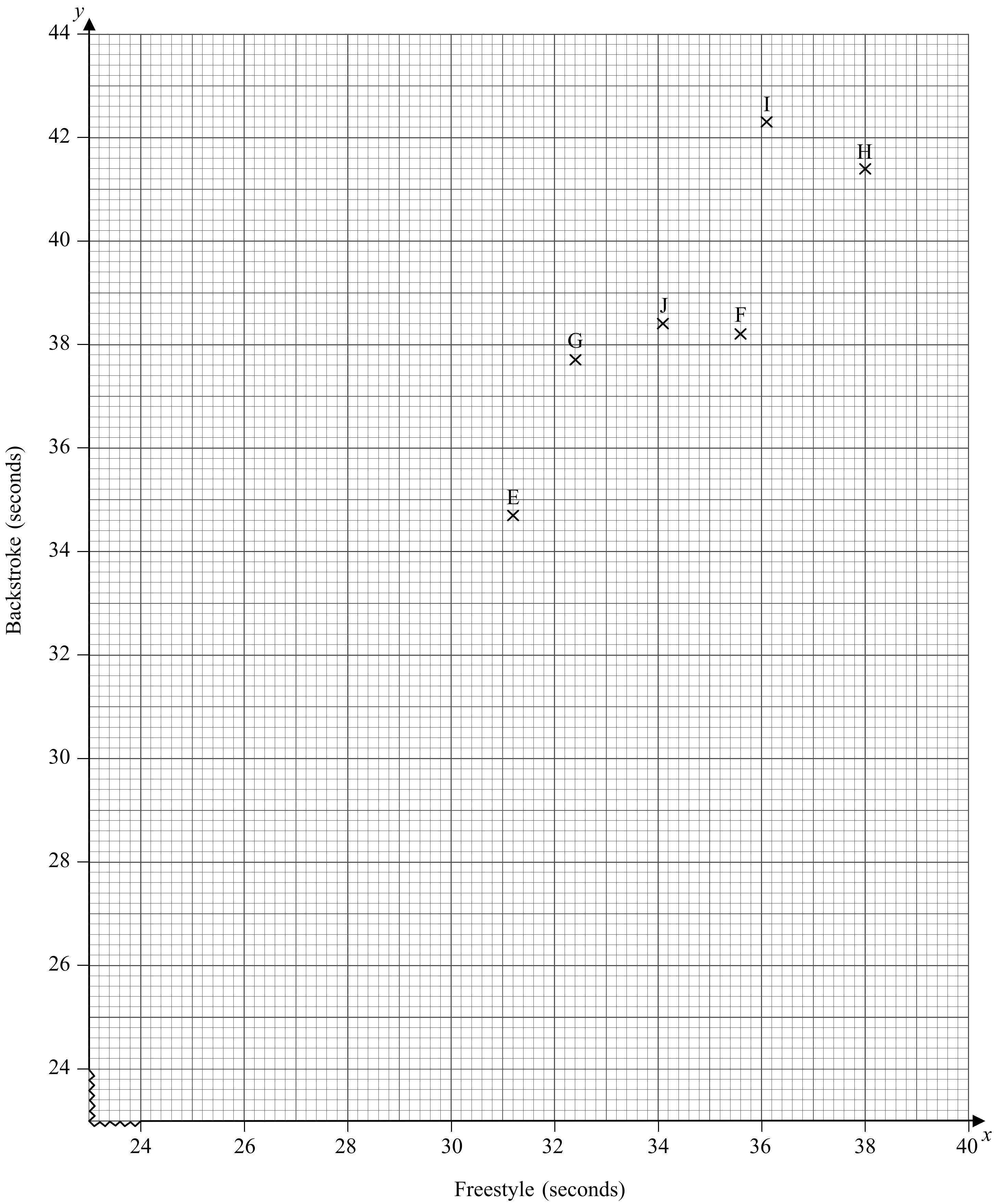
$$310 - \mu = 1.96\sigma$$ *(3 marks)*

(ii) Given that $P(Y < 307.5) = 0.86$, find, to two decimal places, values for μ and σ. *(4 marks)*

END OF QUESTIONS

Figure 1 (for use in Question 5)

Scatter Diagram for Freestyle and Backstroke Swimming Times



Leave blank

1. A personnel manager wants to find out if a test carried out during an employee's interview and a skills assessment at the end of basic training is a guide to performance after working for the company for one year.

The table below shows the results of the interview test of 10 employees and their performance after one year.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Interview test, x %.	65	71	79	77	85	78	85	90	81	62
Performance after one year, y %.	65	74	82	64	87	78	61	65	79	69

[You may use $\sum x^2 = 60\ 475, \sum y^2 = 53\ 122, \sum xy = 56\ 076$]

(a) Showing your working clearly, calculate the product moment correlation coefficient between the interview test and the performance after one year.
(5)

The product moment correlation coefficient between the skills assessment and the performance after one year is –0.156 to 3 significant figures.

(b) Use your answer to part (a) to comment on whether or not the interview test and skills assessment are a guide to the performance after one year. Give clear reasons for your answers.
(2)

Leave blank

2. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724\ 961$]

(a) Find the mean and standard deviation of the level of cotinine in a patient's blood.
(4)

(b) Find the median, upper and lower quartiles of these data.
(3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3+1.5(Q_3-Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

(c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.
(4)

Research suggests that cotinine levels in the blood form a skewed distribution.

One measure of skewness is found using $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$.

(d) Evaluate this measure and describe the skewness of these data.
(3)

Leave blank

3. The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

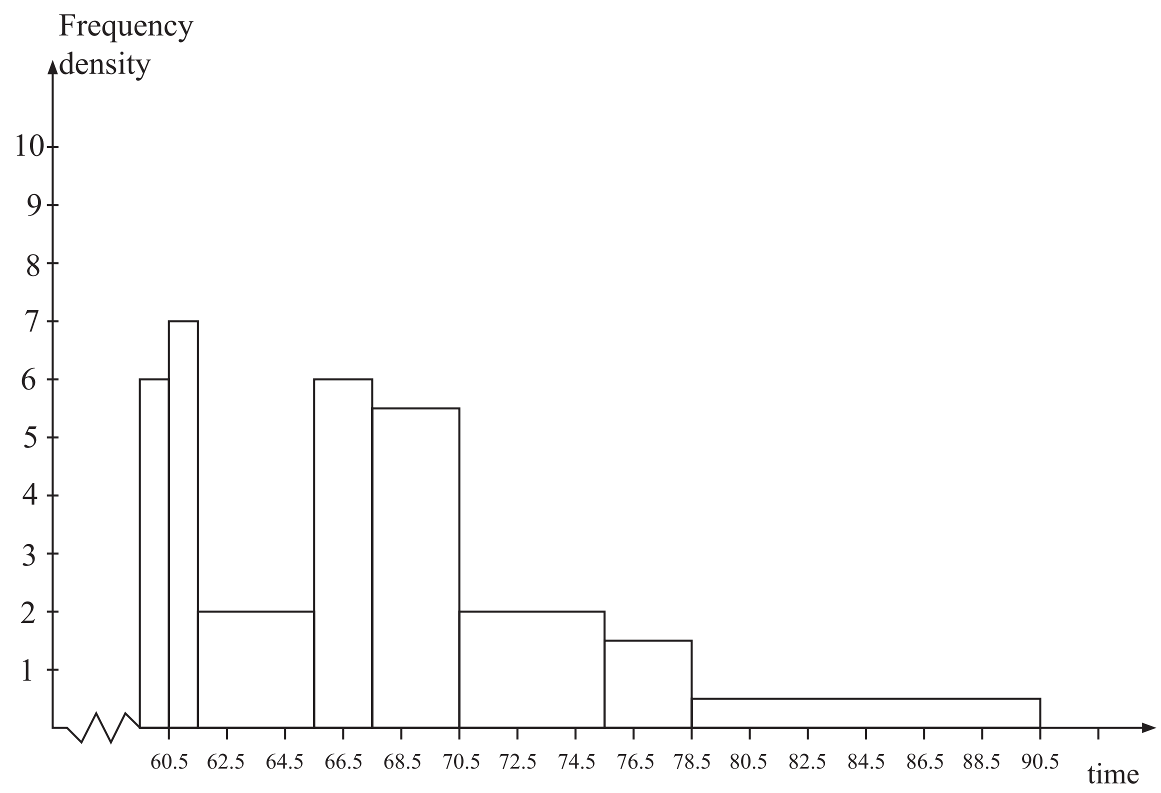


Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(5)

Leave blank

4. A second hand car dealer has 10 cars for sale. She decides to investigate the link between the age of the cars, x years, and the mileage, y thousand miles. The data collected from the cars are shown in the table below.

Age, x (years)	2	2.5	3	4	4.5	4.5	5	3	6	6.5
Mileage, y (thousands)	22	34	33	37	40	45	49	30	58	58

[You may assume that $\sum x = 41, \sum y = 406, \sum x^2 = 188, \sum xy = 1818.5$]

(a) Find S_{xx} and S_{xy}.
(3)

(b) Find the equation of the least squares regression line in the form $y=a+bx$. Give the values of a and b to 2 decimal places.
(4)

(c) Give a practical interpretation of the slope b.
(1)

(d) Using your answer to part (b), find the mileage predicted by the regression line for a 5 year old car.
(2)

Leave blank

5. The following shows the results of a wine tasting survey of 100 people.

96 like wine *A*,
93 like wine *B*,
96 like wine *C*,
92 like *A* and *B*,
91 like *B* and *C*,
93 like *A* and *C*,
90 like all three wines.

(a) Draw a Venn Diagram to represent these data.
(6)

Find the probability that a randomly selected person from the survey likes

(b) none of the three wines,
(1)

(c) wine *A* but not wine *B*,
(2)

(d) any wine in the survey except wine *C*,
(2)

(e) exactly two of the three kinds of wine.
(2)

Given that a person from the survey likes wine *A*,

(f) find the probability that the person likes wine *C*.
(3)

Leave blank

6. The weights of bags of popcorn are normally distributed with mean of 200 g and 60% of all bags weighing between 190 g and 210 g.

(a) Write down the median weight of the bags of popcorn.

(1)

(b) Find the standard deviation of the weights of the bags of popcorn.

(5)

A shopkeeper finds that customers will complain if their bag of popcorn weighs less than 180 g.

(c) Find the probability that a customer will complain.

(3)

Leave blank

7. Tetrahedral dice have four faces. Two fair tetrahedral dice, one red and one blue, have faces numbered 0, 1, 2, and 3 respectively. The dice are rolled and the numbers face down on the two dice are recorded. The random variable R is the score on the red die and the random variable B is the score on the blue die.

(a) Find P(R=3 and B=0).

(2)

The random variable T is R multiplied by B.

(b) Complete the diagram below to represent the sample space that shows all the possible values of T.

3				
2		2		
1	0			
0				
B / *R*	**0**	**1**	**2**	**3**

Sample space diagram of *T*

(3)

(c) The table below represents the probability distribution of the random variable T.

t	0	1	2	3	4	6	9
P(T=t)	a	b	1/8	1/8	c	1/8	d

Find the values of a, b, c and d.

(3)

Find the values of

(d) E(T),

(2)

(e) Var(T).

(4)

1 **(i)** The letters A, B, C, D and E are arranged in a straight line.

(a) How many different arrangements are possible? [2]

(b) In how many of these arrangements are the letters A and B next to each other? [3]

(ii) From the letters A, B, C, D and E, two different letters are selected at random. Find the probability that these two letters are A and B. [2]

2 A random variable T has the distribution $\mathrm{Geo}\left(\frac{1}{5}\right)$. Find

(i) $\mathrm{P}(T = 4)$, [2]

(ii) $\mathrm{P}(T > 4)$, [2]

(iii) $\mathrm{E}(T)$. [1]

3 A sample of bivariate data was taken and the results were summarised as follows.

$$n = 5 \qquad \Sigma x = 24 \qquad \Sigma x^2 = 130 \qquad \Sigma y = 39 \qquad \Sigma y^2 = 361 \qquad \Sigma xy = 212$$

(i) Show that the value of the product moment correlation coefficient r is 0.855, correct to 3 significant figures. [2]

(ii) The ranks of the data were found. One student calculated Spearman's rank correlation coefficient r_s, and found that $r_s = 0.7$. Another student calculated the product moment coefficient, R, of these ranks. State which one of the following statements is true, and explain your answer briefly.

(A) $R = 0.855$
(B) $R = 0.7$
(C) It is impossible to give the value of R without carrying out a calculation using the original data.

[2]

(iii) All the values of x are now multiplied by a scaling factor of 2. State the new values of r and r_s. [2]

4 A supermarket has a large stock of eggs. 40% of the stock are from a firm called Eggzact. 12% of the stock are brown eggs from Eggzact.

An egg is chosen at random from the stock. Calculate the probability that

(i) this egg is brown, given that it is from Eggzact, [2]

(ii) this egg is from Eggzact and is not brown. [2]

5 **(i)** 20% of people in the large town of Carnley support the Residents' Party. 12 people from Carnley are selected at random. Out of these 12 people, the number who support the Residents' Party is denoted by U.

Find

(a) $\mathrm{P}(U \leqslant 5)$, [2]

(b) $\mathrm{P}(U \geqslant 3)$. [3]

(ii) 30% of people in Carnley support the Commerce Party. 15 people from Carnley are selected at random. Out of these 15 people, the number who support the Commerce Party is denoted by V.

Find $\mathrm{P}(V = 4)$. [3]

6 The probability distribution for a random variable Y is shown in the table.

y	1	2	3
$\mathrm{P}(Y = y)$	0.2	0.3	0.5

(i) Calculate $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$. [5]

Another random variable, Z, is independent of Y. The probability distribution for Z is shown in the table.

z	1	2	3
$\mathrm{P}(Z = z)$	0.1	0.25	0.65

One value of Y and one value of Z are chosen at random. Find the probability that

(ii) $Y + Z = 3$, [3]

(iii) $Y \times Z$ is even. [3]

7 **(i)** Andrew plays 10 tennis matches. In each match he either wins or loses.

(a) State, in this context, two conditions needed for a binomial distribution to arise. [2]

(b) Assuming these conditions are satisfied, define a variable in this context which has a binomial distribution. [1]

(ii) The random variable X has the distribution $\mathrm{B}(21, p)$, where $0 < p < 1$.

Given that $\mathrm{P}(X = 10) = \mathrm{P}(X = 9)$, find the value of p. [5]

[Turn over

8 The stem-and-leaf diagram shows the age in completed years of the members of a sports club.

Male		**Female**
8 8 7 6	1	6 6 6 7 7 8 8 9
7 6 5 5 3 3 2 1	2	1 3 3 4 5 7 8 8 9 9
9 8 4 4 3	3	2 3 3 4 7
5 2 1	4	0 1 8
9 0	5	0

Key: 1 | 4 | 0 represents a male aged 41 and a female aged 40.

(i) Find the median and interquartile range for the males. [3]

(ii) The median and interquartile range for the females are 27 and 15 respectively. Make two comparisons between the ages of the males and the ages of the females. [2]

(iii) The mean age of the males is 30.7 and the mean age of the females is 27.5, each correct to 1 decimal place. Give one advantage of using the median rather than the mean to compare the ages of the males with the ages of the females. [1]

A record was kept of the number of hours, X, spent by each member at the club in a year. The results were summarised by

$$n = 49, \quad \Sigma(x - 200) = 245, \quad \Sigma(x - 200)^2 = 9849.$$

(iv) Calculate the mean and standard deviation of X. [6]

9 It is thought that the pH value of sand (a measure of the sand's acidity) may affect the extent to which a particular species of plant will grow in that sand. A botanist wished to determine whether there was any correlation between the pH value of the sand on certain sand dunes, and the amount of each of two plant species growing there. She chose random sections of equal area on each of eight sand dunes and measured the pH values. She then measured the area within each section that was covered by each of the two species. The results were as follows.

	Dune	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
	pH value, x	8.5	8.5	9.5	8.5	6.5	7.5	8.5	9.0
Area, $y\,\text{cm}^2$, covered	Species P	150	150	575	330	45	15	340	330
	Species Q	170	15	80	230	75	25	0	0

The results for species P can be summarised by

$$n = 8, \quad \Sigma x = 66.5, \quad \Sigma x^2 = 558.75, \quad \Sigma y = 1935, \quad \Sigma y^2 = 711\,275, \quad \Sigma xy = 17\,082.5.$$

(i) Give a reason why it might be appropriate to calculate the equation of the regression line of y on x rather than x on y in this situation. [1]

(ii) Calculate the equation of the regression line of y on x for species P, in the form $y = a + bx$, giving the values of a and b correct to 3 significant figures. [4]

(iii) Estimate the value of y for species P on sand where the pH value is 7.0. [2]

The values of the product moment correlation coefficient between x and y for species P and Q are $r_P = 0.828$ and $r_Q = 0.0302$.

(iv) Describe the relationship between the area covered by species Q and the pH value. [1]

(v) State, with a reason, whether the regression line of y on x for species P will provide a reliable estimate of the value of y when the pH value is

(a) 8, [1]

(b) 4. [1]

(vi) Assume that the equation of the regression line of y on x for species Q is also known. State, with a reason, whether this line will provide a reliable estimate of the value of y when the pH value is 8. [1]

Answer **all** questions.

1 The times, in seconds, taken by 20 people to solve a simple numerical puzzle were

17	19	22	26	28	31	34	36	38	39
41	42	43	47	50	51	53	55	57	58

(a) Calculate the mean and the standard deviation of these times. *(3 marks)*

(b) In fact, 23 people solved the puzzle. However, 3 of them failed to solve it within the allotted time of 60 seconds.

Calculate the median and the interquartile range of the times taken by all 23 people. *(4 marks)*

(c) For the times taken by all 23 people, explain why:

(i) the mode is **not** an appropriate numerical measure;

(ii) the range is **not** an appropriate numerical measure. *(2 marks)*

2 The post office in a market town is located within a small supermarket.

The probability that an individual customer entering the supermarket requires a service from:

- the post office only is 0.48;
- the supermarket only is 0.30;
- both the post office and the supermarket is 0.22 .

It may be assumed that the service required is independent from customer to customer.

(a) For a random sample of 12 individual customers, calculate the probability that exactly 5 of them require a service from the post office only. *(3 marks)*

(b) For a random sample of 40 individual customers, determine the probability that more than 10 but fewer than 15 of them require a service from the supermarket only. *(3 marks)*

(c) For a random sample of 100 individual customers, calculate the mean and the standard deviation for the number of them requiring a service from both the post office and the supermarket. *(3 marks)*

3 A very popular play has been performed at a London theatre on each of 6 evenings per week for about a year. Over the past 13 weeks (78 performances), records have been kept of the proceeds from the sales of programmes at each performance. An analysis of these records has found that the mean was £184 and the standard deviation was £32.

(a) Assuming that the 78 performances may be considered to be a random sample, construct a 90% confidence interval for the mean proceeds from the sales of programmes at an evening performance of this play. *(4 marks)*

(b) Comment on the likely validity of the assumption in part (a) when constructing a confidence interval for the mean proceeds from the sales of programmes at an evening performance of:

(i) this particular play;

(ii) any play. *(3 marks)*

4 Rea, Suki and Tora take part in a shooting competition. The final round of the competition requires each of them to try to hit the centre of a target, placed at 100 metres, with a single shot. The independent probabilities that Rea, Suki and Tora hit the centre of this target with a single shot are 0.7, 0.6 and 0.8 respectively.

Find the probability that, in the final round of the competition, the centre of the target will be hit by:

(a) Tora only; *(2 marks)*

(b) exactly one of the three competitors; *(3 marks)*

(c) at least one of the three competitors. *(3 marks)*

Turn over for the next question

5 When Monica walks to work from home, she uses either route A or route B.

(a) Her journey time, X minutes, by route A may be assumed to be normally distributed with a mean of 37 and a standard deviation of 8.

Determine:

(i) $\mathrm{P}(X < 45)$; *(3 marks)*

(ii) $\mathrm{P}(30 < X < 45)$. *(3 marks)*

(b) Her journey time, Y minutes, by route B may be assumed to be normally distributed with a mean of 40 and a standard deviation of σ.

Given that $\mathrm{P}(Y > 45) = 0.12$, calculate the value of σ. *(4 marks)*

(c) If Monica leaves home at 8.15 am to walk to work hoping to arrive by 9.00 am, state, with a reason, which route she should take. *(2 marks)*

6 [**Figure 1**, printed on the insert, is provided for use in this question.]

Stan is a retired academic who supplements his pension by mowing lawns for customers who live nearby.

As part of a review of his charges for this work, he measures the areas, x m^2, of a random sample of eight of his customers' lawns and notes the times, y minutes, that it takes him to mow these lawns. His results are shown in the table.

Customer	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
x	360	140	860	600	1180	540	260	480
y	50	25	135	70	140	90	55	70

(a) On **Figure 1**, plot a scatter diagram of these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. Draw your line on **Figure 1**. *(6 marks)*

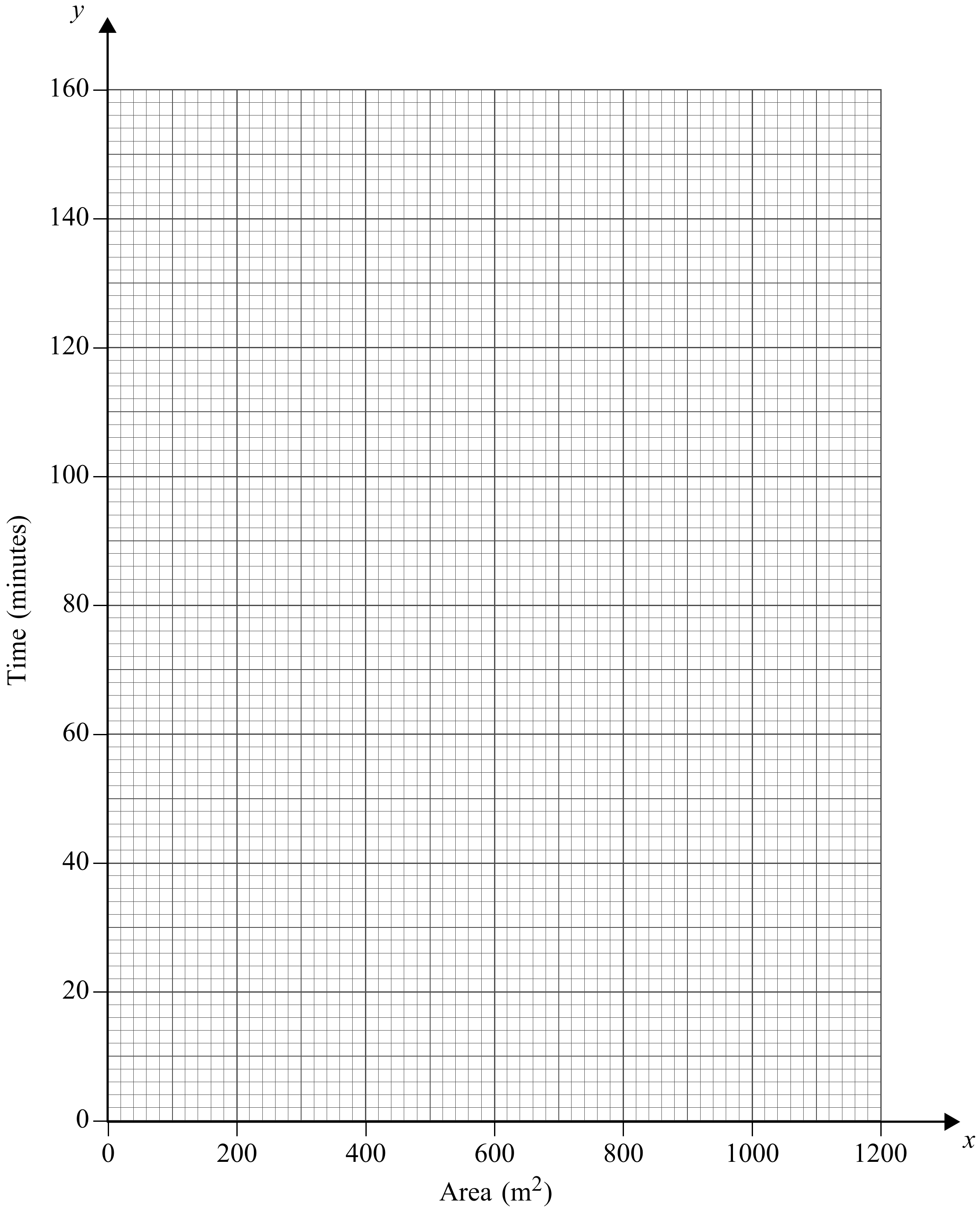
(c) Calculate the value of the residual for Customer H and indicate how your value is confirmed by your scatter diagram. *(3 marks)*

(d) Given that Stan charges £12 per hour, estimate the charge for mowing a customer's lawn that has an area of 560 m^2. *(4 marks)*

END OF QUESTIONS

Figure 1 (for use in Question 6)

Lawn Areas and Mowing Times



Answer **all** questions.

1 The times, in seconds, taken by 20 people to solve a simple numerical puzzle were

17	19	22	26	28	31	34	36	38	39
41	42	43	47	50	51	53	55	57	58

(a) Calculate the mean and the standard deviation of these times. *(3 marks)*

(b) In fact, 23 people solved the puzzle. However, 3 of them failed to solve it within the allotted time of 60 seconds.

Calculate the median and the interquartile range of the times taken by all 23 people. *(4 marks)*

(c) For the times taken by all 23 people, explain why:

(i) the mode is **not** an appropriate numerical measure;

(ii) the range is **not** an appropriate numerical measure. *(2 marks)*

2 A hotel has 50 single rooms, 16 of which are on the ground floor. The hotel offers guests a choice of a full English breakfast, a continental breakfast or no breakfast. The probabilities of these choices being made are 0.45, 0.25 and 0.30 respectively. It may be assumed that the choice of breakfast is independent from guest to guest.

(a) On a particular morning there are 16 guests, each occupying a single room on the ground floor. Calculate the probability that exactly 5 of these guests require a full English breakfast. *(3 marks)*

(b) On a particular morning when there are 50 guests, each occupying a single room, determine the probability that:

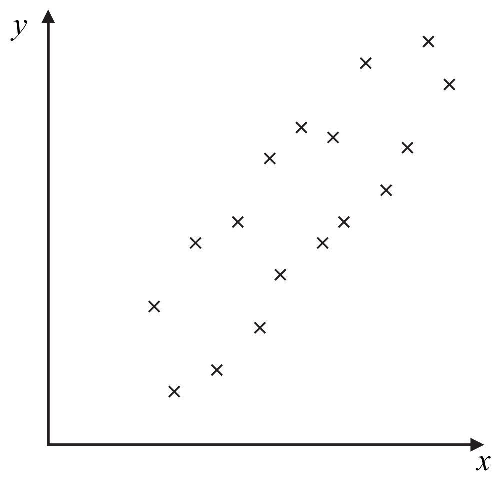
(i) at most 12 of these guests require a continental breakfast; *(2 marks)*

(ii) more than 10 but fewer than 20 of these guests require no breakfast. *(3 marks)*

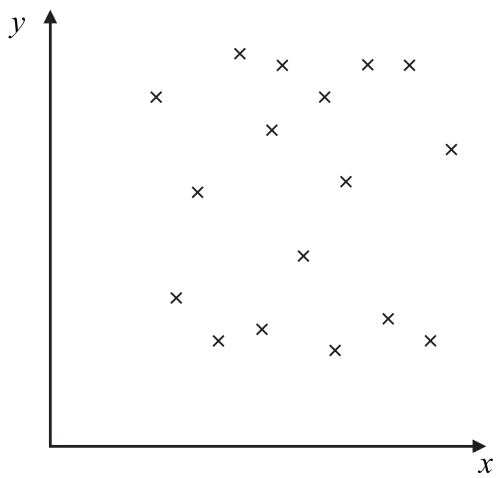
(c) When there are 40 guests, each occupying a single room, calculate the mean and the standard deviation for the number of these guests requiring breakfast. *(4 marks)*

3 Estimate, **without undertaking any calculations**, the value of the product moment correlation coefficient between the variables x and y in each of the three scatter diagrams.

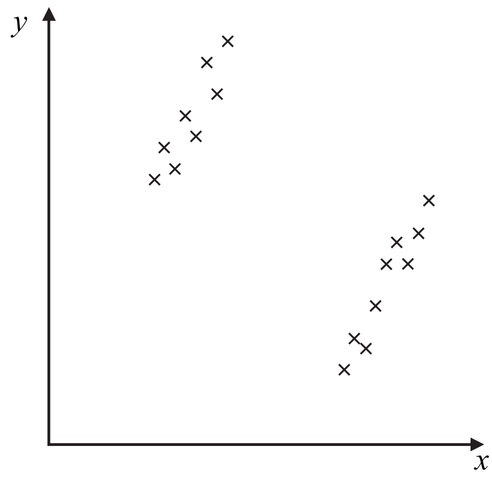
(a)



(b)



(c)



(5 marks)

Turn over for the next question

4 A very popular play has been performed at a London theatre on each of 6 evenings per week for about a year. Over the past 13 weeks (78 performances), records have been kept of the proceeds from the sales of programmes at each performance. An analysis of these records has found that the mean was £184 and the standard deviation was £32.

(a) Assuming that the 78 performances may be considered to be a random sample, construct a 90% confidence interval for the mean proceeds from the sales of programmes at an evening performance of this play. *(4 marks)*

(b) Comment on the likely validity of the assumption in part (a) when constructing a confidence interval for the mean proceeds from the sales of programmes at an evening performance of:

(i) this particular play;

(ii) any play. *(3 marks)*

5 Dafydd, Eli and Fabio are members of an amateur cycling club that holds a time trial each Sunday during the summer. The independent probabilities that Dafydd, Eli and Fabio take part in any one of these trials are 0.6, 0.7 and 0.8 respectively.

Find the probability that, on a particular Sunday during the summer:

(a) none of the three cyclists takes part; *(2 marks)*

(b) Fabio is the only one of the three cyclists to take part; *(2 marks)*

(c) exactly one of the three cyclists takes part; *(3 marks)*

(d) either one or two of the three cyclists take part. *(3 marks)*

6 When Monica walks to work from home, she uses either route A or route B.

(a) Her journey time, X minutes, by route A may be assumed to be normally distributed with a mean of 37 and a standard deviation of 8.

Determine:

(i) $\mathrm{P}(X < 45)$; *(3 marks)*

(ii) $\mathrm{P}(30 < X < 45)$. *(3 marks)*

(b) Her journey time, Y minutes, by route B may be assumed to be normally distributed with a mean of 40 and a standard deviation of σ.

Given that $\mathrm{P}(Y > 45) = 0.12$, calculate the value of σ. *(4 marks)*

(c) If Monica leaves home at 8.15 am to walk to work hoping to arrive by 9.00 am, state, with a reason, which route she should take. *(2 marks)*

(d) When Monica travels to work from home by car, her journey time, W minutes, has a mean of 18 and a standard deviation of 12.

Estimate the probability that, for a random sample of 36 journeys to work from home by car, Monica's mean time is more than 20 minutes. *(4 marks)*

(e) Indicate where, if anywhere, in this question you needed to make use of the Central Limit Theorem. *(1 mark)*

Turn over for the next question

7 [**Figure 1**, printed on the insert, is provided for use in this question.]

Stan is a retired academic who supplements his pension by mowing lawns for customers who live nearby.

As part of a review of his charges for this work, he measures the areas, $x\,\text{m}^2$, of a random sample of eight of his customers' lawns and notes the times, y minutes, that it takes him to mow these lawns. His results are shown in the table.

Customer	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
x	360	140	860	600	1180	540	260	480
y	50	25	135	70	140	90	55	70

(a) On **Figure 1**, plot a scatter diagram of these data. *(2 marks)*

(b) Calculate the equation of the least squares regression line of y on x. Draw your line on **Figure 1**. *(6 marks)*

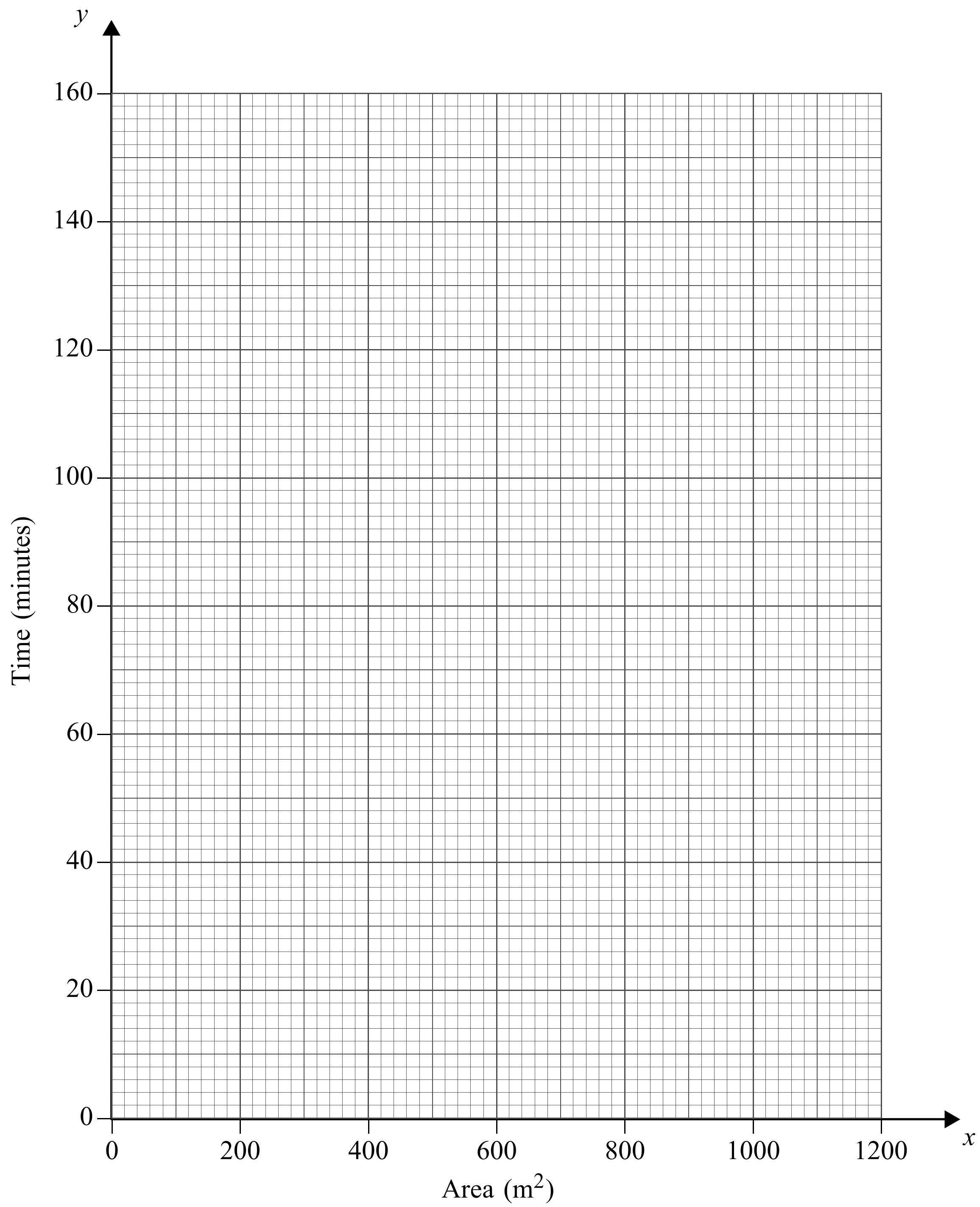
(c) Calculate the value of the residual for Customer H and indicate how your value is confirmed by your scatter diagram. *(3 marks)*

(d) Given that Stan charges £12 per hour, estimate the charge for mowing a customer's lawn that has an area of $560\,\text{m}^2$. *(4 marks)*

END OF QUESTIONS

Figure 1 (for use in Question 7)

Lawn Areas and Mowing Times



Mock papers 7

Leave blank

1. A teacher is monitoring the progress of students using a computer based revision course. The improvement in performance, y marks, is recorded for each student along with the time, x hours, that the student spent using the revision course. The results for a random sample of 10 students are recorded below.

x hours	1.0	3.5	4.0	1.5	1.3	0.5	1.8	2.5	2.3	3.0
y marks	5	30	27	10	–3	–5	7	15	–10	20

[You may use $\sum x = 21.4$, $\sum y = 96$, $\sum x^2 = 57.22$, $\sum xy = 313.7$]

(a) Calculate S_{xx} and S_{xy}.
(3)

(b) Find the equation of the least squares regression line of y on x in the form $y = a + bx$.
(4)

(c) Give an interpretation of the gradient of your regression line.
(1)

Rosemary spends 3.3 hours using the revision course.

(d) Predict her improvement in marks.
(2)

Lee spends 8 hours using the revision course claiming that this should give him an improvement in performance of over 60 marks.

(e) Comment on Lee's claim.
(1)

Leave blank

2. A group of office workers were questioned for a health magazine and $\frac{2}{5}$ were found to take regular exercise. When questioned about their eating habits $\frac{2}{3}$ said they always eat breakfast and, of those who always eat breakfast $\frac{9}{25}$ also took regular exercise.

Find the probability that a randomly selected member of the group

(a) always eats breakfast and takes regular exercise,
(2)

(b) does not always eat breakfast and does not take regular exercise.
(4)

(c) Determine, giving your reason, whether or not always eating breakfast and taking regular exercise are statistically independent.
(2)

Leave blank

3. When Rohit plays a game, the number of points he receives is given by the discrete random variable X with the following probability distribution.

x	0	1	2	3
$P(X = x)$	0.4	0.3	0.2	0.1

(a) Find $E(X)$.
(2)

(b) Find $F(1.5)$.
(2)

(c) Show that $\text{Var}(X) = 1$
(4)

(d) Find $\text{Var}(5 - 3X)$.
(2)

Rohit can win a prize if the total number of points he has scored after 5 games is at least 10. After 3 games he has a total of 6 points.
You may assume that games are independent.

(e) Find the probability that Rohit wins the prize.
(6)

Leave blank

4. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, y minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

(a) Find the median and quartiles for these data. **(3)**

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(b) Show that 110 is the only outlier. **(2)**

(c) Using the graph paper on page 15 draw a box plot for these data indicating clearly the position of the outlier. **(3)**

The value of 110 is omitted.

(d) Show that S_{yy} for the remaining 10 students is 2966.9 **(3)**

These 10 students were each asked how many text messages, x, they sent in the same week.

The values of S_{xx} and S_{xy} for these 10 students are $S_{xx} = 3463.6$ and $S_{xy} = -18.3$.

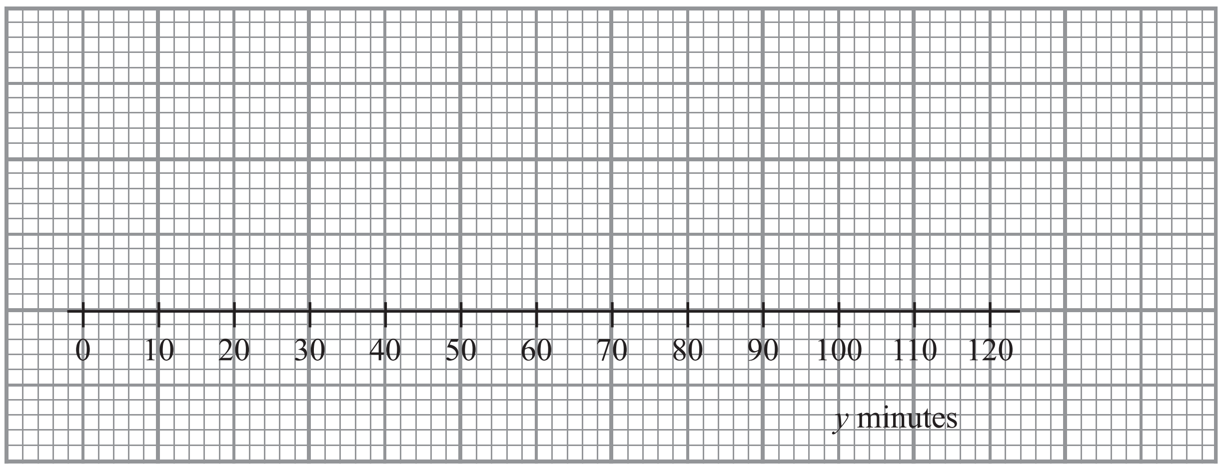
(e) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students. **(2)**

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

(f) Comment on this belief in the light of your calculation in part (e). **(1)**

Leave blank

Question 4 continued



Turn over

Leave blank

5. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

(a) Calculate the width and height of the rectangle representing the group (16 – 25) hours.
(3)

(b) Use linear interpolation to estimate the median and interquartile range.
(5)

(c) Estimate the mean and standard deviation of the number of hours spent shopping.
(4)

(d) State, giving a reason, the skewness of these data.
(2)

(e) State, giving a reason, which average and measure of dispersion you would recommend to use to summarise these data.
(2)

Leave blank

6. The random variable X has a normal distribution with mean 30 and standard deviation 5.

(a) Find $P(X < 39)$.

(2)

(b) Find the value of d such that $P(X < d) = 0.1151$

(4)

(c) Find the value of e such that $P(X > e) = 0.1151$

(2)

(d) Find $P(d < X < e)$.

(2)