

SL Paper 2

As part of his IB Biology field work, Barry was asked to measure the circumference of trees, in centimetres, that were growing at different distances, in metres, from a river bank. His results are summarized in the following table.

Distance, x (metres)	5	12	17	21	24	30	34	44	47
Circumference, y (centimetres)	82	76	70	68	67	60	62	50	50

- a. State whether *distance from the river bank* is a continuous **or** discrete variable. [1]
- b. **On graph paper**, draw a scatter diagram to show Barry's results. Use a scale of 1 cm to represent 5 m on the x -axis and 1 cm to represent 10 cm on the y -axis. [4]
- c. Write down [2]
- (i) the mean distance, \bar{x} , of the trees from the river bank;
 - (ii) the mean circumference, \bar{y} , of the trees.
- d. Plot and label the point $M(\bar{x}, \bar{y})$ on your graph. [2]
- e. Write down [4]
- (i) the Pearson's product-moment correlation coefficient, r , for Barry's results;
 - (ii) the equation of the regression line y on x , for Barry's results.
- f. Draw the regression line y on x on your graph. [2]
- g. **Use the equation of the regression line y on x** to estimate the circumference of a tree that is 40 m from the river bank. [2]

The table below shows the distribution of test grades for 50 IB students at Greendale School.

Test grade	1	2	3	4	5	6	7
Frequency	1	3	7	13	11	10	5

A student is chosen at random from these 50 students.

A second student is chosen at random from these 50 students.

The number of minutes that the 50 students spent preparing for the test was normally distributed with a mean of 105 minutes and a standard deviation of 20 minutes.

- a.i. Calculate the mean test grade of the students; [2]
- a.ii. Calculate the standard deviation. [1]
- b. Find the median test grade of the students. [1]
- c. Find the interquartile range. [2]
- d. Find the probability that this student scored a grade 5 or higher. [2]
- e. Given that the first student chosen at random scored a grade 5 or higher, find the probability that both students scored a grade 6. [3]
- f.i. Calculate the probability that a student chosen at random spent at least 90 minutes preparing for the test. [2]
- f.ii. Calculate the expected number of students that spent at least 90 minutes preparing for the test. [2]

In the month before their IB Diploma examinations, eight male students recorded the number of hours they spent on social media.

For each student, the number of hours spent on social media (x) and the number of IB Diploma points obtained (y) are shown in the following table.

Hours on social media (x)	6	15	26	12	13	40	33	23
IB Diploma points (y)	43	33	27	36	39	17	20	33

Use your graphic display calculator to find

Ten female students also recorded the number of hours they spent on social media in the month before their IB Diploma examinations. Each of these female students spent between 3 and 30 hours on social media.

The equation of the regression line y on x for these ten female students is

$$y = -\frac{2}{3}x + \frac{125}{3}.$$

An eleventh girl spent 34 hours on social media in the month before her IB Diploma examinations.

- a. On graph paper, draw a scatter diagram for these data. Use a scale of 2 cm to represent 5 hours on the x -axis and 2 cm to represent 10 points on the y -axis. [4]
- b. (i) \bar{x} , the mean number of hours spent on social media; [2]
(ii) \bar{y} , the mean number of IB Diploma points.
- c. Plot the point (\bar{x}, \bar{y}) on your scatter diagram and label this point M. [2]
- d. Write down the value of r , the Pearson's product-moment correlation coefficient, for these data. [2]

- e. Write down the equation of the regression line y on x for these eight male students. [2]
- f. Draw the regression line, from part (e), on your scatter diagram. [2]
- g. Use the given equation of the regression line to estimate the number of IB Diploma points that this girl obtained. [2]
- h. Write down a reason why this estimate is not reliable. [1]

A survey was conducted to determine the length of time, t , in minutes, people took to drink their coffee in a café. The information is shown in the following grouped frequency table.

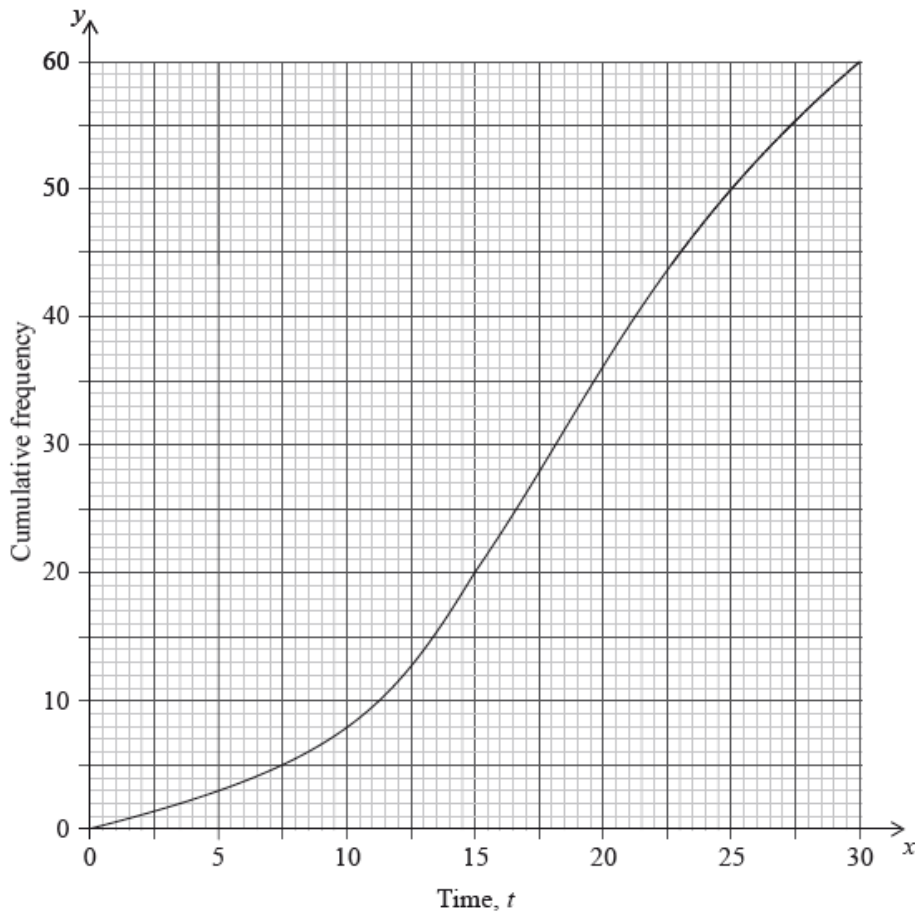
Time, t (minutes)	Number of People
$0 < t \leq 5$	3
$5 < t \leq 10$	5
$10 < t \leq 15$	12
$15 < t \leq 20$	14
$20 < t \leq 25$	16
$25 < t \leq 30$	10

- a. Write down the total number of people who were surveyed. [1]
- b. Write down the mid-interval value for the $10 < t \leq 15$ group. [1]
- c. Find an estimate of the mean time people took to drink their coffee. [2]
- d. The information above has been rewritten as a cumulative frequency table. [2]

Time, t (minutes)	$t \leq 5$	$t \leq 10$	$t \leq 15$	$t \leq 20$	$t \leq 25$	$t \leq 30$
Cumulative frequency	3	8	20	a	50	b

Write down the value of a and the value of b .

- e. This information is shown in the following cumulative frequency graph. [4]



For the people who were surveyed, use the graph to estimate

- (i) the time taken for the first 40 people to drink their coffee;
- (ii) the number of people who take less than 8 minutes to drink their coffee;
- (iii) the number of people who take more than 23 minutes to drink their coffee.

A group of 800 students answered 40 questions on a category of their choice out of History, Science and Literature.

For each student the category and the number of correct answers, N , was recorded. The results obtained are represented in the following table.

		Number of correct answers, N				Total number of students
		$1 \leq N \leq 10$	$11 \leq N \leq 20$	$21 \leq N \leq 30$	$31 \leq N \leq 40$	
Category	History	46	80	68	39	233
	Science	37	82	85	56	260
	Literature	31	110	104	62	307
	Total number of students	114	272	257	157	800

A χ^2 test at the 5% significance level is carried out on the results. The critical value for this test is 12.592.

a. State whether N is a discrete or a continuous variable.

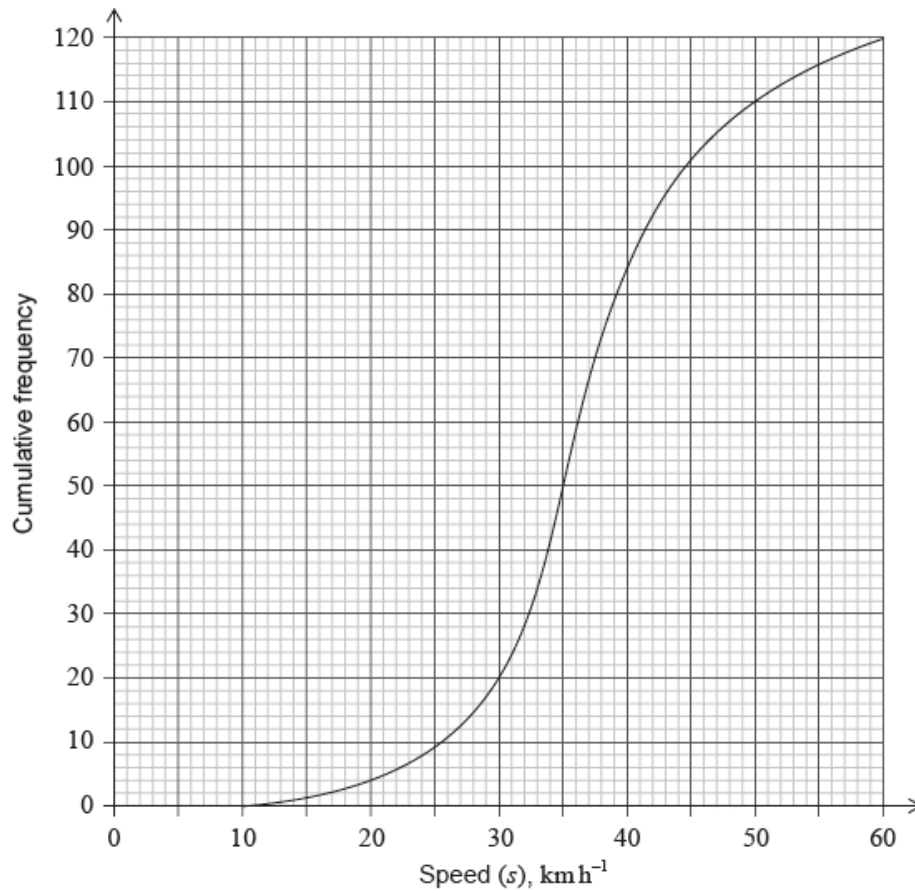
[1]

b.i. Write down, for N , the modal class;

[1]

- b.ii. Write down, for N , the mid-interval value of the modal class. [1]
- c.i. Use your graphic display calculator to estimate the mean of N ; [2]
- c.ii. Use your graphic display calculator to estimate the standard deviation of N . [1]
- d. Find the expected frequency of students choosing the Science category and obtaining 31 to 40 correct answers. [2]
- e.i. Write down the null hypothesis for this test; [1]
- e.ii. Write down the number of degrees of freedom. [1]
- f.i. Write down the p -value for the test; [1]
- f.ii. Write down the χ^2 statistic. [2]
- g. State the result of the test. Give a reason for your answer. [2]

The cumulative frequency graph shows the speed, s , in km h^{-1} , of 120 vehicles passing a hospital gate.



- a. Estimate the minimum possible speed of one of these vehicles passing the hospital gate. [1]
- b. Find the median speed of the vehicles. [2]
- c. Write down the 75th percentile. [1]

d. Calculate the interquartile range. [2]

e. The speed limit past the hospital gate is 50 km h^{-1} . [2]

Find the number of these vehicles that exceed the speed limit.

f. The table shows the speeds of these vehicles travelling past the hospital gate. [2]

Speed of Vehicles	Number of Vehicles
$0 < s \leq 10$	0
$10 < s \leq 20$	p
$20 < s \leq 30$	16
$30 < s \leq 40$	64
$40 < s \leq 50$	26
$50 < s \leq 60$	q

Find the value of p and of q .

g. The table shows the speeds of these vehicles travelling past the hospital gate. [2]

Speed of Vehicles	Number of Vehicles
$0 < s \leq 10$	0
$10 < s \leq 20$	p
$20 < s \leq 30$	16
$30 < s \leq 40$	64
$40 < s \leq 50$	26
$50 < s \leq 60$	q

(i) Write down the modal class.

(ii) Write down the mid-interval value for this class.

h. The table shows the speeds of these vehicles travelling past the hospital gate. [3]

Speed of Vehicles	Number of Vehicles
$0 < s \leq 10$	0
$10 < s \leq 20$	p
$20 < s \leq 30$	16
$30 < s \leq 40$	64
$40 < s \leq 50$	26
$50 < s \leq 60$	q

Use your graphic display calculator to calculate an estimate of

(i) the mean speed of these vehicles;

(ii) the standard deviation.

i. It is proposed that the speed limit past the hospital gate is reduced to 40 km h^{-1} from the current 50 km h^{-1} . [2]

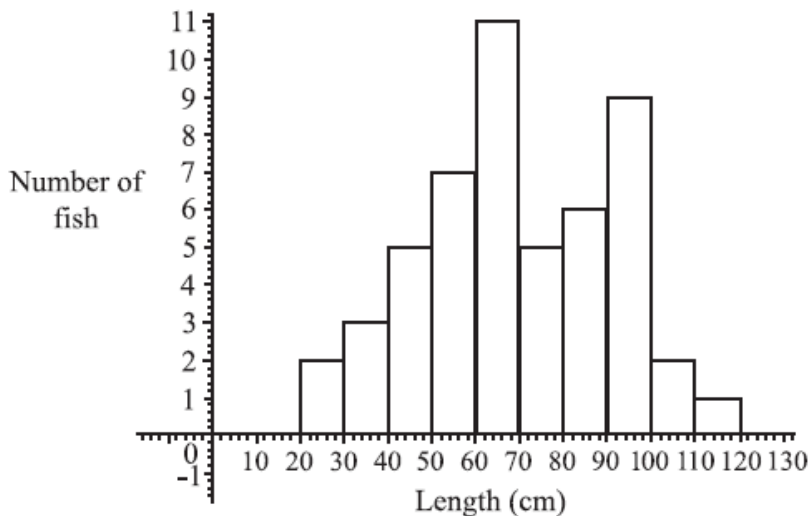
Find the percentage of these vehicles passing the hospital gate that **do not** exceed the current speed limit but **would** exceed the new speed limit.

The table below shows the scores for 12 golfers for their first two rounds in a local golf tournament.

Round 1 (x)	71	79	66	73	69	76	68	75	82	67	69	74
Round 2 (y)	73	81	68	75	70	79	69	77	83	68	72	76

- a. (i) Write down the mean score in Round 1. [5]
- (ii) Write down the standard deviation in Round 1.
- (iii) Find the number of these golfers that had a score of more than one standard deviation above the mean in Round 1. [2]
- b. Write down the correlation coefficient, r . [2]
- c. Write down the equation of the regression line of y on x . [2]
- d. Another golfer scored 70 in Round 1. [2]
- Calculate an estimate of his score in Round 2.
- e. Another golfer scored 89 in Round 1. [2]
- Determine whether you can use the equation of the regression line to estimate his score in Round 2. Give a reason for your answer.

The figure below shows the lengths in centimetres of fish found in the net of a small trawler.



- a. Find the total number of fish in the net. [2]
- b. Find (i) the modal length interval, [5]
- (ii) the interval containing the median length,
- (iii) an estimate of the mean length.

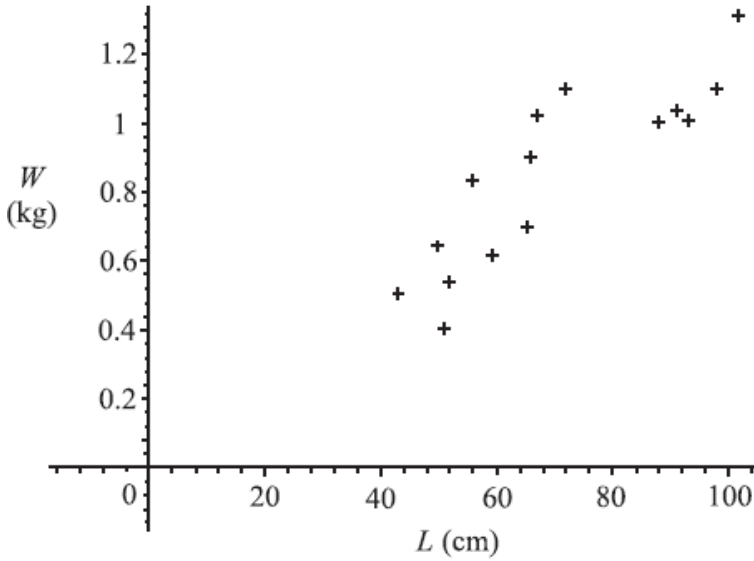
c. (i) Write down an estimate for the standard deviation of the lengths. [3]

(ii) How many fish (if any) have length **greater than** three standard deviations **above** the mean?

d. The fishing company must pay a fine if more than 10% of the catch have lengths less than 40cm. [2]

Do a calculation to decide whether the company is fined.

e. A sample of 15 of the fish was weighed. The weight, W was plotted against length, L as shown below. [2]



Exactly **two** of the following statements about the plot could be correct. Identify the two correct statements.

Note: You do **not** need to enter data in a GDC **or** to calculate r exactly.

(i) The value of r , the correlation coefficient, is approximately 0.871.

(ii) There is an exact linear relation between W and L .

(iii) The line of regression of W on L has equation $W = 0.012L + 0.008$.

(iv) There is negative correlation between the length and weight.

(v) The value of r , the correlation coefficient, is approximately 0.998.

(vi) The line of regression of W on L has equation $W = 63.5L + 16.5$.

Daniel grows apples and chooses at random a sample of 100 apples from his harvest.

He measures the diameters of the apples to the nearest cm. The following table shows the distribution of the diameters.

Diameter (to the nearest cm)	5	6	7	8	9
Frequency	15	27	33	17	8

a. Using your graphic display calculator, write down the value of [3]

(i) the mean of the diameters in this sample;

(ii) the standard deviation of the diameters in this sample.

b. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of [3]

1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

Calculate the percentage of **small** apples in Daniel's harvest.

- c. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of [2] 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

Of the apples harvested, 5% are **large** apples.

Find the value of a .

- d. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of [2] 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

Find the percentage of **medium** apples.

- e. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of [2] 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

This year, Daniel estimates that he will grow 100 000 apples.

Estimate the number of **large** apples that Daniel will grow this year.

The table shows the distance, in km, of eight regional railway stations from a city centre terminus and the price, in \$, of a return ticket from each regional station to the terminus.

Distance in km (x)	3	15	23	42	56	62	74	93
Price in \$ (y)	5	24	43	56	68	74	86	100

- a. Draw a scatter diagram for the above data. Use a scale of 1 cm to represent 10 km on the x -axis and 1 cm to represent \$10 on the y -axis. [4]
- b. Use your graphic display calculator to find [2]
- (i) \bar{x} , the mean of the distances;
- (ii) \bar{y} , the mean of the prices.
- c. Plot and label the point M (\bar{x} , \bar{y}) on your scatter diagram. [1]
- d. Use your graphic display calculator to find [3]
- (i) the product-moment correlation coefficient, r ;
- (ii) the equation of the regression line y on x .
- e. Draw the regression line y on x on your scatter diagram. [2]
- f. A ninth regional station is 76 km from the city centre terminus. [3]
- Use the equation of the regression line to estimate the price of a return ticket to the city centre terminus from this regional station. **Give your answer correct to the nearest \$.**
- g. Give a reason why it is valid to use your regression line to estimate the price of this return ticket. [1]
- h. The actual price of the return ticket is \$80. [2]
- Using your answer to part (f),** calculate the percentage error in the estimated price of the ticket.

The lengths (l) in centimetres of 100 copper pipes at a local building supplier were measured. The results are listed in the table below.

Length l (cm)	Frequency
17.5	12
32.5	26
47.5	32
62.5	21
77.5	9

- a. Write down the mode. [1]

- b. Using your graphic display calculator, write down the value of [4]
- (i) the mean;
 - (ii) the standard deviation;
 - (iii) the median.
- c. Find the interquartile range. [2]
- d. Draw a box and whisker diagram for this data, on graph paper, using a scale of 1 cm to represent 5 cm. [4]
- e. Sam estimated the value of the mean of the measured lengths to be 43 cm. [2]
- Find the percentage error of Sam's estimated mean.

Alex and Kris are riding their bicycles together along a bicycle trail and note the following distance markers at the given times.

Time (t hours)	1	2	3	4	5	6	7
Distance (d km)	57	65	72	81	89	97	107

- a. Draw a scatter diagram of the data. Use 1 cm to represent 1 hour and 1 cm to represent 10 km. [3]
- b.i. Write down for this set of data the mean time, \bar{t} . [1]
- b.ii. Write down for this set of data the mean distance, \bar{d} . [1]
- c. Mark and label the point $M(\bar{t}, \bar{d})$ on your scatter diagram. [2]
- d. Draw the line of best fit on your scatter diagram. [2]
- e. **Using your graph**, estimate the time when Alex and Kris pass the 85 km distance marker. Give your answer correct to **one decimal place**. [2]
- f. Write down the equation of the regression line for the data given. [2]
- g.i. **Using your equation** calculate the distance marker passed by the cyclists at 10.3 hours. [2]
- g.ii. Is this estimate of the distance reliable? Give a reason for your answer. [2]

In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured. The results are given in the table below.

Number of trees (x)	Depth of snow in cm (y)
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

In a study on 100 students there seemed to be a difference between males and females in their choice of favourite car colour. The results are given in the table below. A χ^2 test was conducted.

	Blue	Red	Green
Males	14	6	8
Females	31	24	17

A, ~~Use~~ your graphic display calculator to find the mean number of trees. [1]

A, ~~Use~~ your graphic display calculator to find the mean depth of snow. [1]

A, ~~Use~~ your graphic display calculator to find the standard deviation of the depth of snow. [1]

A, ~~The~~ covariance, $S_{xy} = 188.5$. [2]

Write down the product-moment correlation coefficient, r .

A, ~~Write~~ down the equation of the regression line of y on x . [2]

A, ~~if~~ the number of trees in an area is 55, estimate the depth of snow. [2]

A, ~~Use~~ the equation of the regression line to estimate the depth of snow in an area with 100 trees. [1]

A, ~~Decide~~ whether the answer in (e)(i) is a valid estimate of the depth of snow in the area. Give a reason for your answer. [2]

B, ~~Write~~ down the total number of male students. [1]

B, ~~Show~~ that the expected frequency for males, whose favourite car colour is blue, is 12.6. [2]

B, ~~The~~ calculated value of χ^2 is 1.367 and the critical value of χ^2 is 5.99 at the 5% significance level. [1]

Write down the null hypothesis for this test.

B, ~~The~~ calculated value of χ^2 is 1.367 and the critical value of χ^2 is 5.99 at the 5% significance level. [1]

Write down the number of degrees of freedom.

B, ~~The~~ calculated value of χ^2 is 1.367 and the critical value of χ^2 is 5.99 at the 5% significance level. [2]

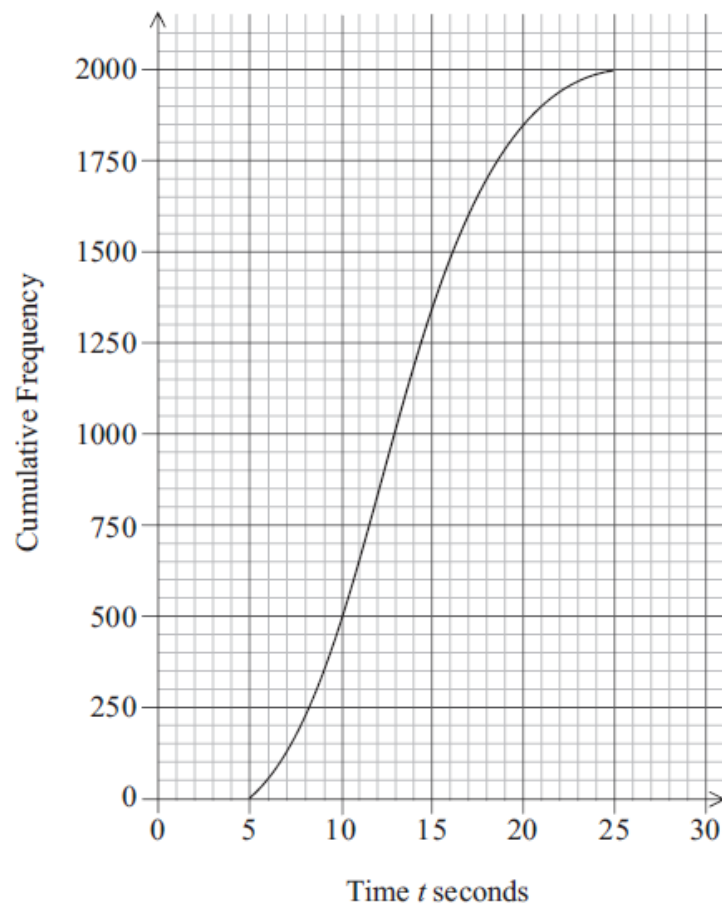
Determine whether the null hypothesis should be accepted at the 5% significance level. Give a reason for your answer.

The number of bottles of water sold at a railway station on each day is given in the following table.

Day	0	1	2	3	4	5	6	7	8	9	10	11	12
Temperature (T°)	21	20.7	20	19	18	17.3	17	17.3	18	19	20	20.7	21
Number of bottles sold (n)	150	141	126	125	98	101	93	99	116	121	119	134	141

- a. Write down [2]
- (i) the mean temperature;
 - (ii) the standard deviation of the temperatures.
- b. Write down the correlation coefficient, r , for the variables n and T . [1]
- c. Comment on your value for r . [2]
- d. The equation of the line of regression for n on T is $n = dT - 100$. [2]
- (i) Write down the value of d .
 - (ii) Estimate how many bottles of water will be sold when the temperature is 19.6° .
- e. On a day when the temperature was 36° Peter calculates that 314 bottles would be sold. Give one reason why his answer might be unreliable. [1]

The diagram shows the cumulative frequency graph for the time t taken to perform a certain task by 2000 men.



a, i Use the diagram to estimate the median time.

[1]

a, ii Use the diagram to estimate the upper quartile and the lower quartile.

[2]

a, iii Use the diagram to estimate the interquartile range.

[1]

b. Find the number of men who take **more than** 11 seconds to perform the task.

[3]

c. 55 % of the men took less than p seconds to perform the task. Find p .

[2]

d, i The times taken for the 2000 men were grouped as shown in the table below.

[1]

Time	Frequency
$5 \leq t < 10$	500
$10 \leq t < 15$	850
$15 \leq t < 20$	a
$20 \leq t < 25$	b

Write down the value of a .

d, ii The times taken for the 2000 men were grouped as shown in the table below.

[1]

 onbekend.png

Write down the value of b .

e, i Use your graphic display calculator to find an estimate of the mean time. [2]

e, ii Use your graphic display calculator to find an estimate of the standard deviation of the time. [1]

f. Everyone who performs the task in **less than** one standard deviation **below** the mean will receive a bonus. Pedro takes 9.5 seconds to perform the task. [3]

Does Pedro receive the bonus? Justify your answer.

On one day 180 flights arrived at a particular airport. The distance travelled and the arrival status for each incoming flight was recorded. The flight was then classified as on time, slightly delayed, or heavily delayed.

The results are shown in the following table.

		Distance travelled			TOTAL
		At most 500 km	Between 500 km and 5000 km	At least 5000 km	
Arrival Status	On time	19	17	16	52
	Slightly delayed	13	18	14	45
	Heavily delayed	28	15	40	83
TOTAL		60	50	70	180

A χ^2 test is carried out at the 10 % significance level to determine whether the arrival status of incoming flights is independent of the distance travelled.

The critical value for this test is 7.779.

A flight is chosen at random from the 180 recorded flights.

a. State the alternative hypothesis. [1]

b. Calculate the expected frequency of flights travelling at most 500 km and arriving slightly delayed. [2]

c. Write down the number of degrees of freedom. [1]

d.i. Write down the χ^2 statistic. [2]

d.ii. Write down the associated p -value. [1]

e. State, with a reason, whether you would reject the null hypothesis. [2]

f. Write down the probability that this flight arrived on time. [2]

g. Given that this flight was not heavily delayed, find the probability that it travelled between 500 km and 5000 km. [2]

h. Two flights are chosen at random from those which were slightly delayed.

[3]

Find the probability that each of these flights travelled at least 5000 km.

A random sample of 167 people who own mobile phones was used to collect data on the amount of time they spent per day using their phones.

The results are displayed in the table below.

Time spent per day (t minutes)	$0 \leq t < 15$	$15 \leq t < 30$	$30 \leq t < 45$	$45 \leq t < 60$	$60 \leq t < 75$	$75 \leq t < 90$
Number of people	21	32	35	41	27	11

Manuel conducts a survey on a random sample of 751 people to see which television programme type they watch most from the following:

Drama, Comedy, Film, News. The results are as follows.

	Drama	Comedy	Film	News
Males under 25	22	65	90	35
Males 25 and over	36	54	67	17
Females under 25	22	59	82	15
Females 25 and over	64	39	38	46

Manuel decides to ignore the ages and to test at the 5 % level of significance whether the most watched programme type is independent of **gender**.

i.a. State the modal group.

[1]

i.b. Use your graphic display calculator to calculate approximate values of the mean and standard deviation of the time spent per day on these mobile phones.

[3]

i.c. On graph paper, draw a fully labelled histogram to represent the data.

[4]

ii.a. Draw a table with 2 rows and 4 columns of data so that Manuel can perform a chi-squared test.

[3]

ii.b. State Manuel's null hypothesis and alternative hypothesis.

[1]

ii.c. Find the expected frequency for the number of females who had 'Comedy' as their most-watched programme type. Give your answer to the nearest whole number.

[2]

ii.d. Using your graphic display calculator, or otherwise, find the chi-squared statistic for Manuel's data.

[3]

ii.e.(i) State the number of degrees of freedom available for this calculation.

[3]

(ii) State his conclusion.

The speed, s , in km h^{-1} , of 120 vehicles passing a point on the road was measured. The results are given below.

Speed, s (km h^{-1})	$0 < s \leq 50$	$50 < s \leq 60$	$60 < s \leq 70$	$70 < s \leq 80$	$80 < s \leq 90$	$90 < s \leq 100$
Number of vehicles	30	46	22	12	8	2

a. Write down the midpoint of the $60 < s \leq 70$ interval. [1]

b. Use your graphic display calculator to find an estimate for [3]

- (i) the mean speed of the vehicles;
- (ii) the standard deviation of the speeds of the vehicles.

c. Write down the number of vehicles whose speed is less than or equal to 60 km h^{-1} . [1]

d. Consider the cumulative frequency table below. [2]

Speed, s (km h^{-1})	$s \leq 50$	$s \leq 60$	$s \leq 70$	$s \leq 80$	$s \leq 90$	$s \leq 100$
Number of vehicles	30	a	b	110	c	120

Write down the value of a , of b and of c .

e. Consider the cumulative frequency table below. [4]

Speed, s (km h^{-1})	$s \leq 50$	$s \leq 60$	$s \leq 70$	$s \leq 80$	$s \leq 90$	$s \leq 100$
Number of vehicles	30	a	b	110	c	120

Draw a cumulative frequency graph for the information from the table. Use 1 cm to represent 10 km h^{-1} on the horizontal axis and 1 cm to represent 10 vehicles on the vertical axis.

f. Use your cumulative frequency graph to estimate [4]

- (i) the median speed of the vehicles;
- (ii) the number of vehicles that are travelling at a speed less than or equal to 65 km h^{-1} .

g. All drivers whose vehicle's speed is greater than one standard deviation above the speed limit of 50 km h^{-1} will be fined. [3]

Use your graph to estimate the number of drivers who will be fined.

Francesca is a chef in a restaurant. She cooks eight chickens and records their masses and cooking times. The mass m of each chicken, in kg, and its cooking time t , in minutes, are shown in the following table.

Mass m (kg)	Cooking time t (minutes)
1.5	62
1.6	75
1.8	82
1.9	83
2.0	86
2.1	87
2.1	91
2.3	98

- a. Draw a scatter diagram to show the relationship between the mass of a chicken and its cooking time. Use 2 cm to represent 0.5 kg on the horizontal axis and 1 cm to represent 10 minutes on the vertical axis. [4]
- b. Write down for this set of data [2]
- (i) the mean mass, \bar{m} ;
- (ii) the mean cooking time, \bar{t} .
- c. Label the point $M(\bar{m}, \bar{t})$ on the scatter diagram. [1]
- d. Draw the line of best fit on the scatter diagram. [2]
- e. Using your line of best fit, estimate the cooking time, in minutes, for a 1.7 kg chicken. [2]
- f. Write down the Pearson's product-moment correlation coefficient, r . [2]
- g. Using your value for r , comment on the correlation. [2]
- h. The cooking time of an additional 2.0 kg chicken is recorded. If the mass and cooking time of this chicken is included in the data, the correlation [2]
- is weak.
- (i) Explain how the cooking time of this additional chicken might differ from that of the other eight chickens.
- (ii) Explain how a new line of best fit might differ from that drawn in part (d).

Part A

A university required all Science students to study one language for one year. A survey was carried out at the university amongst the 150 Science students. These students all studied one of either French, Spanish or Russian. The results of the survey are shown below.

	French	Spanish	Russian
Female	9	29	12
Male	31	40	29

Ludmila decides to use the χ^2 test at the 5% level of significance to determine whether the choice of language is independent of gender.

At the end of the year, only seven of the female Science students sat examinations in Science and French. The marks for these seven students are shown in the following table.

Science (S)	23	51	56	62	12	73	72
French (F)	65	45	45	40	70	36	30

A.a State Ludmila's null hypothesis. [1]

A.b Write down the number of degrees of freedom. [1]

A.c Find the expected frequency for the females studying Spanish. [2]

A.d Use your graphic display calculator to find the χ^2 test statistic for this data. [2]

A.e State whether Ludmila accepts the null hypothesis. Give a reason for your answer. [2]

B.a Draw a labelled scatter diagram for this data. Use a scale of 2 cm to represent 10 marks on the x -axis (S) and 10 marks on the y -axis (F). [4]

B.b Use your graphic calculator to find [2]

(i) \bar{S} , the mean of S ;

(ii) \bar{F} , the mean of F .

B.c Plot the point $M(\bar{S}, \bar{F})$ on your scatter diagram. [1]

B.d Use your graphic display calculator to find the equation of the regression line of F on S . [2]

B.e Draw the regression line on your scatter diagram. [2]

B.f Carletta's mark on the Science examination was 44. She did not sit the French examination. [2]

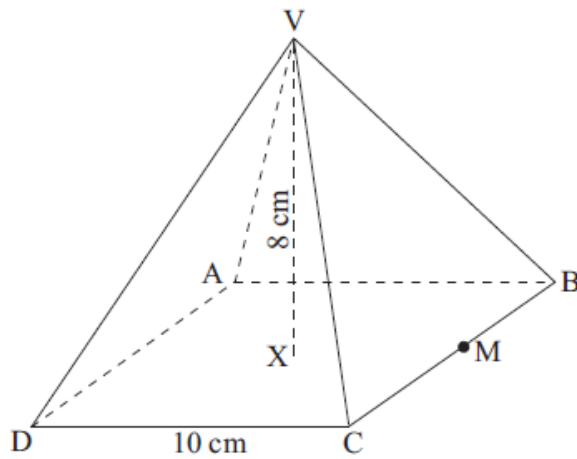
Estimate Carletta's mark for the French examination.

B.g Monique's mark on the Science examination was 85. She did not sit the French examination. Her French teacher wants to use the regression line to estimate Monique's mark. [2]

State whether the mark obtained from the regression line for Monique's French examination is reliable. Justify your answer.

The diagram below shows a square based right pyramid. ABCD is a square of side 10 cm. VX is the perpendicular height of 8 cm. M is the midpoint of BC.

diagram not to scale



In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured. The results are given in the table below.

Number of trees (x)	Depth of snow in cm (y)
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

A path goes around a forest so that it forms the three sides of a triangle. The lengths of two sides are 550 m and 290 m. These two sides meet at an angle of 115° . A diagram is shown below.

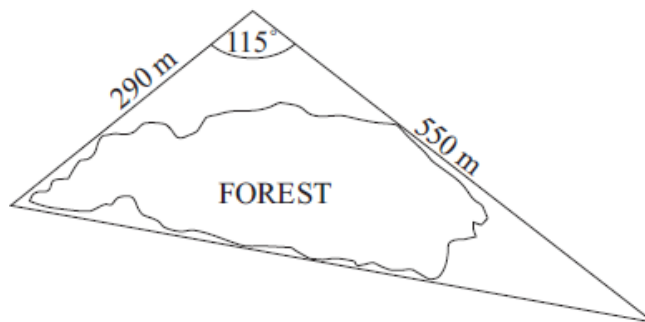


diagram not to scale

A, write down the length of XM.

[1]

A, use your graphic display calculator to find the standard deviation of the number of trees.

[1]

A, calculate the length of VM.

[2]

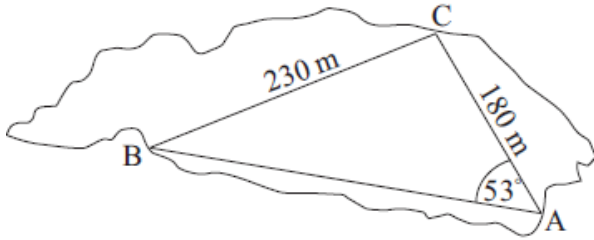
A, Calculate the angle between VM and ABCD. [2]

B, Calculate the length of the third side of the triangle. Give your answer correct to the nearest 10 m. [4]

B, Calculate the area enclosed by the path that goes around the forest. [3]

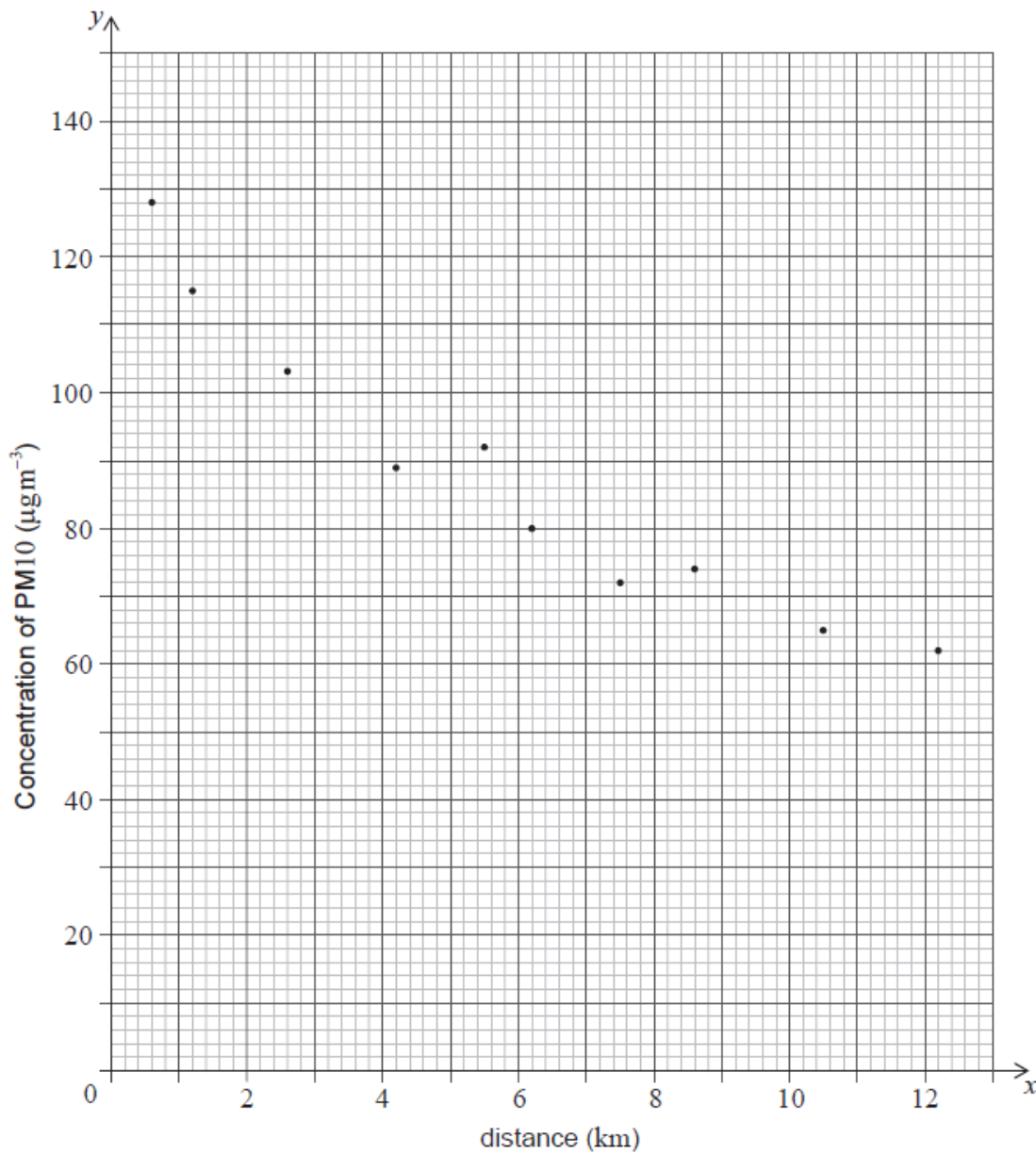
B, Inside the forest a second path forms the three sides of another triangle named ABC. Angle BAC is 53° , AC is 180 m and BC is 230 m. [4]

diagram not to scale



Calculate the size of angle ACB.

a. For an ecological study, Ernesto measured the average concentration (y) of the fine dust, PM10, in the air at different distances (x) from a power plant. His data are represented on the following scatter diagram. The concentration of PM10 is measured in micrograms per cubic metre and the distance is measured in kilometres. [2]



His data are also listed in the following table.

Distance (x)	0.6	1.2	2.6	a	5.5	6.2	7.5	8.6	10.5	12.2
Concentration of PM10 (y)	128	115	103	89	92	80	72	b	65	62

Use the scatter diagram to find the value of a and of b in the table.

b. Calculate

[4]

- i) \bar{x} , the mean distance from the power plant;
- ii) \bar{y} , the mean concentration of PM10 ;
- iii) r , the Pearson's product-moment correlation coefficient.

c. Write down the equation of the regression line y on x .

[2]

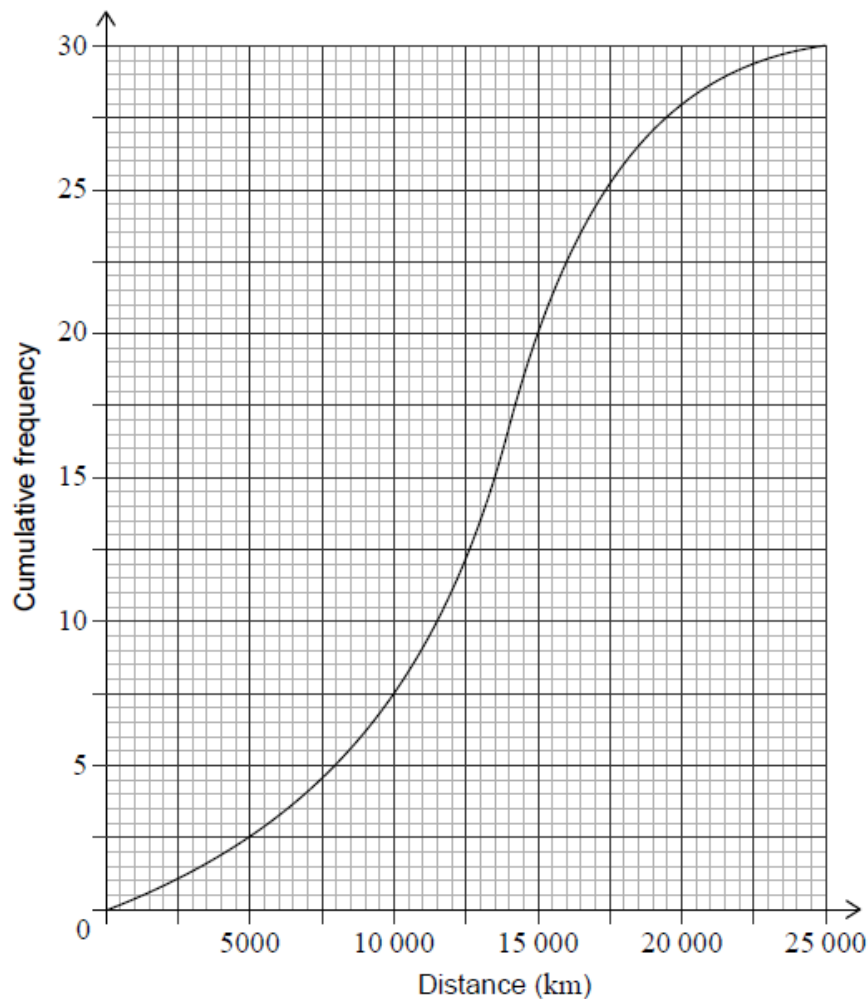
d. Ernesto's school is located 14 km from the power plant. He uses the equation of the regression line to estimate the concentration of PM10 in the air at his school.

[4]

- i) Calculate the value of Ernesto's estimate.

ii) State whether Ernesto's estimate is reliable. Justify your answer.

A transportation company owns 30 buses. The distance that each bus has travelled since being purchased by the company is recorded. The cumulative frequency curve for these data is shown.



It is known that 8 buses travelled more than m kilometres.

- a. Find the number of buses that travelled a distance between 15000 and 20000 kilometres. [2]
- b.i. Use the cumulative frequency curve to find the median distance. [2]
- b.ii. Use the cumulative frequency curve to find the lower quartile. [1]
- b.iii. Use the cumulative frequency curve to find the upper quartile. [1]
- c. Hence write down the interquartile range. [1]
- d. Write down the percentage of buses that travelled a distance greater than the upper quartile. [1]
- e. Find the number of buses that travelled a distance less than or equal to 12000 km. [1]

f. Find the value of m . [2]

g. The smallest distance travelled by one of the buses was 2500 km. [4]

The longest distance travelled by one of the buses was 23 000 km.

On graph paper, draw a box-and-whisker diagram for these data. Use a scale of 2 cm to represent 5000 km.

The heat output in thermal units from burning 1 kg of wood changes according to the wood's percentage moisture content. The moisture content and heat output of 10 blocks of the same type of wood each weighing 1 kg were measured. These are shown in the table.

Moisture content % (x)	8	15	22	30	34	45	50	60	74	82
Heat output (y)	80	77	74	69	68	61	61	55	50	45

a. Draw a scatter diagram to show the above data. Use a scale of 2 cm to represent 10% on the x -axis and a scale of 2 cm to represent 10 thermal units on the y -axis. [4]

b. Write down [2]

(i) the mean percentage moisture content, \bar{x} ;

(ii) the mean heat output, \bar{y} .

c. Plot the point (\bar{x}, \bar{y}) on your scatter diagram and label this point M . [2]

d. Write down the product-moment correlation coefficient, r . [2]

e. The equation of the regression line y on x is $y = -0.470x + 83.7$. Draw the regression line y on x on your scatter diagram. [2]

f. The equation of the regression line y on x is $y = -0.470x + 83.7$. Estimate the heat output in thermal units of a 1 kg block of wood that has 25% moisture content. [2]

g. The equation of the regression line y on x is $y = -0.470x + 83.7$. State, with a reason, whether it is appropriate to use the regression line y on x to estimate the heat output in part (f). [2]

In an environmental study of plant diversity around a lake, a biologist collected data about the number of different plant species (y) that were growing at different distances (x) in metres from the lake shore.

Distance (x)	2	5	8	10	13	17	23	35	40
Plant species (y)	35	34	30	29	24	19	15	13	8

- a. Draw a scatter diagram to show the data. Use a scale of 2 cm to represent 10 metres on the x -axis and 2 cm to represent 10 plant species on the y -axis. [4]
- b. Using your scatter diagram, describe the correlation between the number of different plant species and the distance from the lake shore. [1]
- c.i. Use your graphic display calculator to write down \bar{x} , the mean of the distances from the lake shore. [1]
- c.ii. Use your graphic display calculator to write down \bar{y} , the mean number of plant species. [1]
- d. Plot the point (\bar{x}, \bar{y}) on your scatter diagram. **Label this point M.** [2]
- e. Write down the equation of the regression line y on x for the above data. [2]
- f. Draw the regression line y on x on your scatter diagram. [2]
- g. Estimate the number of plant species growing 30 metres from the lake shore. [2]

200 people were asked the amount of time T (minutes) they had spent in the supermarket. The results are represented in the table below.

Time (T)	$0 < T \leq 10$	$10 < T \leq 20$	$20 < T \leq 30$	$30 < T \leq 40$	$40 < T \leq 50$
Number of people	23	57	93	21	6

- a. State if the data is discrete or continuous. [1]
- b. State the modal group. [1]
- c. Write down the midpoint of the interval $10 < T \leq 20$. [1]
- d. Use your graphic display calculator to find an estimate for [3]
- (i) the mean;
- (ii) the standard deviation.
- e. The results are represented in the cumulative frequency table below, with upper class boundaries of 10, 20, 30, 40, 50. [2]

Upper class boundaries	10	20	30	40	50
Cumulative frequency	23	80	173	q	r

Write down the value of

- (i) q ;
- (ii) r .
- f. The results are represented in the cumulative frequency table below, with upper class boundaries of 10, 20, 30, 40, 50. [4]

Upper class boundaries	10	20	30	40	50
Cumulative frequency	23	80	173	q	r

On graph paper, draw a cumulative frequency graph, using a scale of 2 cm to represent 10 minutes (T) on the horizontal axis and 1 cm to represent 10 people on the vertical axis.

g. **Use your graph** from part (f) to estimate

[6]

(i) the median;

(ii) the 90th percentile of the results;

(iii) the number of people who shopped at the supermarket for more than 15 minutes.
