SL Paper 2

The manager of a folder factory recorded the number of folders produced by the factory (in thousands) and the production costs (in thousand Euros), for six consecutive months.

	January	February	March	April	May	June
Folders produced, x (thousands)	10	15	22	30	28	21
Production costs, C (thousand Euros)	35	40	60	72	68	55

Every month the factory sells all the folders produced. Each folder is sold for 2.99 Euros.

a. Draw a scatter diagram for this data. Use a scale of 2 cm for 5000 folders on the horizontal axis and 2 cm for 10 000 Euros on the vertical axis. [4]

b.i. Write down, for this set of data the mean number of folders produced, $\bar{x}$; [1]

b.ii. Write down, for this set of data the mean production cost, $\bar{C}$. [1]

c. Label the point $M(\bar{x}, \ \bar{C})$ on the scatter diagram. [1]

d. Use your graphic display calculator to find the Pearson's product–moment correlation coefficient, r. [2]

e. State a reason why the regression line C on x is appropriate to model the relationship between these variables. [1]

f. Use your graphic display calculator to find the equation of the regression line C on x. [2]

g. Draw the regression line C on x on the scatter diagram. [2]

h. Use the equation of the regression line to estimate the least number of folders that the factory needs to sell in a month to exceed its production cost for that month. [4]

A manufacturer produces 1500 boxes of breakfast cereal every day.

The weights of these boxes are normally distributed with a mean of 502 grams and a standard deviation of 2 grams.

All boxes of cereal with a weight between 497.5 grams and 505 grams are sold. The manufacturer's income from the sale of each box of cereal is $2.00.

The manufacturer recycles any box of cereal with a weight **not** between 497.5 grams and 505 grams. The manufacturer's recycling cost is $0.16 per box.

A **different** manufacturer produces boxes of cereal with weights that are normally distributed with a mean of 350 grams and a standard deviation of 1.8 grams.

This manufacturer sells all boxes of cereal that are above a minimum weight, w.

They sell 97% of the cereal boxes produced.

a. Draw a diagram that shows this information. [2]

b. (i)   Find the probability that a box of cereal, chosen at random, is sold. [4]

   (ii)   Calculate the manufacturer's expected daily income from these sales.

c. Calculate the manufacturer's expected daily recycling cost. [2]

d. Calculate the value of w. [3]

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 following table shows the number of bicycles, x, produced daily by a factory and their total production cost, y, in US dollars (USD). The table shows data recorded over seven days.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Number of bicycles, x	12	15	14	17	20	18	21
Production cost, y	3900	4600	4100	5300	6000	5400	6000

a. (i) Write down the Pearson's product–moment correlation coefficient, r, for these data.   [4]

(ii) Hence comment on the result.

b. Write down the equation of the regression line y on x for these data, in the form $y = ax + b$.   [2]

c. Estimate the total cost, **to the nearest USD**, of producing 13 bicycles on a particular day.   [3]

d. All the bicycles that are produced are sold. The bicycles are sold for 304 USD **each**.   [2]

Explain why the factory does **not** make a profit when producing 13 bicycles on a particular day.

e. All the bicycles that are produced are sold. The bicycles are sold for 304 USD **each**.   [5]

(i) Write down an expression for the total selling price of x bicycles.

(ii) Write down an expression for the **profit** the factory makes when producing x bicycles on a particular day.

(iii) Find the least number of bicycles that the factory should produce, on a particular day, in order to make a profit.

The following table shows the average body weight, x, and the average weight of the brain, y, of seven species of mammal. Both measured in kilograms (kg).

Species	Average body weight, x (kg)	Average weight of the brain, y (kg)
Cat	3	0.026
Cow	465	0.423
Donkey	187	0.419
Giraffe	529	0.680
Goat	28	0.115
Jaguar	100	0.157
Sheep	56	0.175

The average body weight of grey wolves is 36 kg.

In fact, the average weight of the brain of grey wolves is 0.120 kg.

The average body weight of mice is 0.023 kg.

a. Find the range of the average body weights for these seven species of mammal. [2]

b.i.For the data from these seven species calculate r, the Pearson's product–moment correlation coefficient; [2]

b.iiFor the data from these seven species describe the correlation between the average body weight and the average weight of the brain. [2]

c. Write down the equation of the regression line y on x, in the form $y = mx + c$. [2]

d. Use your regression line to estimate the average weight of the brain of grey wolves. [2]

e. Find the percentage error in your estimate in part (d). [2]

f. State whether it is valid to use the regression line to estimate the average weight of the brain of mice. Give a reason for your answer. [2]

A group of candidates sat a Chemistry examination and a Physics examination. The candidates' marks in the Chemistry examination are normally distributed with a mean of 60 and a standard deviation of 12.

a. Draw a diagram that shows this information. [2]

b. Write down the probability that a randomly chosen candidate who sat the Chemistry examination scored at most 60 marks. [1]

c. Hee Jin scored 80 marks in the Chemistry examination. [2]

Find the probability that a randomly chosen candidate who sat the Chemistry examination scored **more** than Hee Jin.

d. The candidates' marks in the Physics examination are normally distributed with a mean of 63 and a standard deviation of 10. Hee Jin also [2]

scored 80 marks in the Physics examination.

Find the probability that a randomly chosen candidate who sat the Physics examination scored **less** than Hee Jin.

e. The candidates' marks in the Physics examination are normally distributed with a mean of 63 and a standard deviation of 10. Hee Jin also [2]

scored 80 marks in the Physics examination.

Determine whether Hee Jin's Physics mark, **compared to the other candidates**, is better than her mark in Chemistry. Give a reason for your answer.

f. To obtain a "grade A" a candidate must be in the top 10% of the candidates who sat the Physics examination. [3]

Find the minimum possible mark to obtain a "grade A". Give your answer correct to the nearest integer.

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 a school, all Mathematical Studies SL students were given a test. The test contained four questions, each one on a different topic from the syllabus. The quality of each response was classified as satisfactory or not satisfactory. Each student answered only three of the four questions, each on a separate answer sheet.

The table below shows the number of satisfactory and not satisfactory responses for each question.

		Topic from the syllabus				
		Calculus	Probability	Geometry	Logic	Total
Quality of response	Satisfactory	10	16	20	14	60
	Not satisfactory	8	6	10	6	30
	Total	18	22	30	20	90

A χ^2 test is carried out at the 5% significance level for the data in the table.

The critical value for this test is 7.815.

a.i. If the teacher chooses a response at random, find the probability that it is a response to the Calculus question; [2]

a.ii. If the teacher chooses a response at random, find the probability that it is a satisfactory response to the Calculus question; [2]

a.iii. If the teacher chooses a response at random, find the probability that it is a satisfactory response, given that it is a response to the Calculus [2]

question.

b. The teacher groups the responses by topic, and chooses two responses to the Logic question. Find the probability that both are not [3]

satisfactory.

c. State the null hypothesis for this test. [1]

d. Show that the expected frequency of satisfactory Calculus responses is 12. [1]

e. Write down the number of degrees of freedom for this test. [1]

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

g. State the conclusion of this χ^2 test. Give a reason for your answer. [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 [4]

on the y-axis.

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]

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.

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.

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

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

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

A, b)The covariance, S_{xy} = 188.5. [2]

Write down the product-moment correlation coefficient, r.

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

A, d)If the number of trees in an area is 55, estimate the depth of snow. [2]

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

A, e)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, a)Write down the total number of male students. [1]

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

B, c)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, c)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, c)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.

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				
		$1 \le N \le 10$	$11 \le N \le 20$	$21 \le N \le 30$	$31 \le N \le 40$	Total number of students
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]

Part A

100 students are asked what they had for breakfast on a particular morning. There were three choices: cereal (X) , bread (Y) and fruit (Z). It is found that

    10 students had all three

    17 students had bread and fruit only

    15 students had cereal and fruit only

12 students had cereal and bread only

13 students had only bread

8 students had only cereal

9 students had only fruit

Part B

The same 100 students are also asked how many meals on average they have per day. The data collected is organized in the following table.

	3 or fewer meals per day	4 or 5 meals per day	More than 5 meals per day	Total
Male	15	25	15	55
Female	12	20	13	45
Total	27	45	28	100

A χ^2 test is carried out at the 5 % level of significance.

A.a Represent this information on a Venn diagram. [4]

A.b Find the number of students who had none of the three choices for breakfast. [2]

A.c Write down the percentage of students who had fruit for breakfast. [2]

A.d Describe in words what the students in the set $X \cap Y'$ had for breakfast. [2]

A.e Find the probability that a student had **at least** two of the three choices for breakfast. [2]

A.f Two students are chosen at random. Find the probability that both students had all three choices for breakfast. [3]

B.a Write down the null hypothesis, H_0, for this test. [1]

B.b Write down the number of degrees of freedom for this test. [1]

B.c Write down the critical value for this test. [1]

B.d Show that the expected number of females that have more than 5 meals per day is 13, correct to the nearest integer. [2]

B.e Use your graphic display calculator to find the χ^2_{calc} for this data. [2]

B.f Decide whether H_0 must be accepted. Justify your answer. [2]

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.

In a debate on voting, a survey was conducted. The survey asked people's opinion on whether or not the minimum voting age should be reduced to 16 years of age. The results are shown as follows.

	Age 18–25	Age 26–40	Age 41+	Total
Oppose the reduction	12	20	48	80
Favour the reduction	18	15	17	50
Total	30	35	65	130

A χ^2 test at the 1% significance level was conducted. The χ^2 critical value of the test is 9.21.

a. State [2]

   (i) H_0, the null hypothesis for the test;

   (ii) H_1, the alternative hypothesis for the test.

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

c. Show that the expected frequency of those between the ages of 26 and 40 who oppose the reduction in the voting age is 21.5, correct to three [2] significant figures.

d. Find [3]

   (i)   the χ^2 statistic;

   (ii) the associated p-value for the test.

e. Determine, giving a reason, whether H_0 should be accepted. [2]

The seniors from Gulf High School are required to participate in exactly one after-school sport. Data were gathered from a sample of 120 students regarding their choice of sport. The following data were recorded.

Gender	Sport			
	Football	Tennis	Basketball	Total
Male	17	8	10	35
Female	31	17	37	85
Total	48	25	47	120

A χ^2 test was carried out at the 5 % significance level to analyse the relationship between gender and choice of after-school sport.

a. Write down the null hypothesis, H_0, for this test. [1]

b. Find the expected value of female footballers. [2]

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

d. Write down the critical value of χ^2, at the 5 % level of significance. [1]

e. Use your graphic display calculator to determine the χ^2_{calc} value. [2]

f. Determine whether H_0 should be accepted. Justify your answer. [2]

g. One student is chosen at random from the 120 students. [2]

Find the probability that this student

(i) is male;

(ii) plays tennis.

h. Two students are chosen at random from the 120 students. [5]

Find the probability that

(i) both play football;

(ii) neither play basketball.

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 ≤ Diameter < a
Large	Diameter ≥ 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 ≤ Diameter < a
Large	Diameter ≥ 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 ≤ Diameter < a
Large	Diameter ≥ 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	$6.5 \leq$ 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.

An agricultural cooperative uses three brands of fertilizer, A, B and C, on 120 different crops. The crop yields are classified as High, Medium or Low.

The data collected are organized in the table below.

	Fertilizer			Total
	A	B	C	
High Yield	10	8	12	30
Medium Yield	24	14	12	50
Low Yield	16	8	16	40
Total	50	30	40	120

The agricultural cooperative decides to conduct a chi-squared test at the 1 % significance level using the data.

a. State the null hypothesis, H_0, for the test. [2]

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

c. Write down the critical value for the test. [1]

d. Show that the expected number of Medium Yield crops using Fertilizer C is 17, correct to the nearest integer. [2]

e. Use your graphic display calculator to find for the data [3]

   (i) the χ^2 calculated value, χ^2_{calc};

   (ii) the p-value.

f. State the conclusion of the test. Give a reason for your decision. [2]

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]

---

a. A speed camera on Peterson Road records the speed of each passing vehicle. The speeds are found to be normally distributed with a mean of   [2]
$67\,\mathrm{km\,h^{-1}}$ and a standard deviation of $3.4\,\mathrm{km\,h^{-1}}$.

    Sketch a diagram of this normal distribution and shade the region representing the probability that the speed of a vehicle is between 60 and $70\,\mathrm{km\,h^{-1}}$.

b. A vehicle on Peterson Road is chosen at random.   [3]

    Find the probability that the speed of this vehicle is

    (i)   more than $60\,\mathrm{km\,h^{-1}}$;

    (ii)   less than $70\,\mathrm{km\,h^{-1}}$;

    (iii) between 60 and $70\,\mathrm{km\,h^{-1}}$.

c. It is found that $19\,\%$ of the vehicles are exceeding the speed limit of $s\,\mathrm{km\,h^{-1}}$.   [2]

    Find the value of s , correct to the nearest integer.

d. There is a fine of US\$65 for exceeding the speed limit on Peterson Road. On a particular day the total value of fines issued was US\$14 820.   [4]

    (i)   Calculate the number of fines that were issued on this day.

    (ii)   Estimate the total number of vehicles that passed the speed camera on Peterson Road on this day.

One day the numbers of customers at three cafés, "Alan's Diner" (A), "Sarah's Snackbar" (S) and "Pete's Eats" (P), were recorded and are given below.

17 were customers of Pete's Eats only

27 were customers of Sarah's Snackbar only

15 were customers of Alan's Diner only

10 were customers of Pete's Eats **and** Sarah's Snackbar **but not** Alan's Diner

8 were customers of Pete's Eats **and** Alan's Diner **but not** Sarah's Snackbar

Some of the customers in each café were given survey forms to complete to find out if they were satisfied with the standard of service they received.

	Pete's Eats	Alan's Diner	Sarah's Snackbar	Total
Dissatisfied	16	8	16	40
Satisfied	26	20	34	80
Total	42	28	50	120

A.aDraw a Venn Diagram, using sets labelled A, S and P, that shows this information. [3]

A.bThere were 48 customers of Pete's Eats that day. Calculate the number of people who were customers of all three cafés. [2]

A.cThere were 50 customers of Sarah's Snackbar that day. Calculate the total number of people who were customers of Alan's Diner. [3]

A.dWrite down the number of customers of Alan's Diner that were also customers of Pete's Eats. [1]

A.eFind $n[(S \cup P) \cap A']$. [2]

B.aOne of the survey forms was chosen at random, find the probability that the form showed "Dissatisfied"; [2]

B.bOne of the survey forms was chosen at random, find the probability that the form showed "Satisfied" and was completed at Sarah's Snackbar; [2]

B.cOne of the survey forms was chosen at random, find the probability that the form showed "Dissatisfied", given that it was completed at Alan's [2]
Diner.

B.dA χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each [1]
of the cafés.

Write down the null hypothesis, H_0, for the χ^2 test.

B.eA χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each [1]
of the cafés.

Write down the number of degrees of freedom for the test.

B.fA χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each [2]
of the cafés.

Using your graphic display calculator, find χ^2_{calc}.

B.gA χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each [2]
of the cafés.

State, giving a reason, the conclusion to the test.

The Brahma chicken produces eggs with weights in grams that are normally distributed about a mean of 55 g with a standard deviation of 7 g. The eggs are classified as small, medium, large or extra large according to their weight, as shown in the table below.

Size	Weight (g)
Small	Weight < 53
Medium	$53 \leq$ Weight < 63
Large	$63 \leq$ Weight < 73
Extra Large	Weight ≥ 73

a. Sketch a diagram of the distribution of the weight of Brahma chicken eggs. On your diagram, show clearly the boundaries for the classification of the eggs. [3]

b. An egg is chosen at random. Find the probability that the egg is [4]

(i) medium;

(ii) extra large.

c. There is a probability of 0.3 that a randomly chosen egg weighs more than w grams. [2]

Find w .

d. The probability that a Brahma chicken produces a large size egg is 0.121. Frank's Brahma chickens produce 2000 eggs each month. [2]

Calculate an estimate of the number of large size eggs produced by Frank's chickens each month.

e. The selling price, in US dollars (USD), of each size is shown in the table below. [3]

Size	Selling price (USD)
Small	0.30
Medium	0.50
Large	0.65
Extra Large	0.80

The probability that a Brahma chicken produces a small size egg is 0.388.

Estimate the monthly income, in USD, earned by selling the 2000 eggs. Give your answer correct to two decimal places.

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]

A store recorded their sales of televisions during the 2010 football World Cup. They looked at the numbers of televisions bought by gender and the size of the television screens.

This information is shown in the table below; S represents the size of the television screen in inches.

	$S \leq 22$	$22 < S \leq 32$	$32 < S \leq 46$	$S > 46$	Total
Female	65	100	40	15	220
Male	20	65	140	55	280
Total	85	165	180	70	500

The store wants to use this information to predict the probability of selling these sizes of televisions for the 2014 football World Cup.

a. Use the table to find the probability that [6]

(i) a television will be bought by a female;

(ii) a television with a screen size of $32 < S \leq 46$ will be bought;

(iii) a television with a screen size of $32 < S \leq 46$ will be bought by a female;

(iv) a television with a screen size greater than 46 inches will be bought, given that it is bought by a male.

b. The manager of the store wants to determine whether the screen size is independent of gender. A Chi-squared test is performed at the 1 % [1]
significance level.

Write down the null hypothesis.

c. The manager of the store wants to determine whether the screen size is independent of gender. A Chi-squared test is performed at the 1 % [2]

significance level.

Show that the expected frequency for females who bought a screen size of $32 < S \leq 46$, is 79, correct to the nearest integer.

d. The manager of the store wants to determine whether the screen size is independent of gender. A Chi-squared test is performed at the 1 % [1]

significance level.

Write down the number of degrees of freedom.

e. The manager of the store wants to determine whether the screen size is independent of gender. A Chi-squared test is performed at the 1 % [2]

significance level.

Write down the χ^2 calculated value.

f. The manager of the store wants to determine whether the screen size is independent of gender. A Chi-squared test is performed at the 1 % [1]

significance level.

Write down the critical value for this test.

g. The manager of the store wants to determine whether the screen size is independent of gender. A Chi-squared test is performed at the 1 % [2]

significance level.

Determine if the null hypothesis should be accepted. Give a reason for your answer.

Pam has collected data from a group of 400 IB Diploma students about the Mathematics course they studied and the language in which they were

examined (English, Spanish or French). The summary of her data is given below.

	Mathematics HL	Mathematics SL	Mathematical Studies SL	Total
English	50	70	80	200
Spanish	30	50	30	110
French	20	30	40	90
Total	100	150	150	400

a. A student is chosen at random from the group. Find the probability that the student [8]

(i)   studied Mathematics HL;

(ii)   was examined in French;

(iii) studied Mathematics HL and was examined in French;

(iv)   did not study Mathematics SL and was not examined in English;

(v)   studied Mathematical Studies SL given that the student was examined in Spanish.

b. Pam believes that the Mathematics course a student chooses is independent of the language in which the student is examined. [2]

Using your answers to parts (a) (i), (ii) and (iii) above, state whether there is any evidence for Pam's belief. Give a reason for your answer.

c. Pam decides to test her belief using a Chi-squared test at the 5% level of significance. [3]

(i)   State the null hypothesis for this test.

(ii) Show that the expected number of Mathematical Studies SL students who took the examination in Spanish is 41.3, correct to 3 significant figures.

d. Write down [4]

(i)   the Chi-squared calculated value;

(ii) the number of degrees of freedom;

(iii) the Chi-squared critical value.

e. State, giving a reason, whether there is sufficient evidence at the 5% level of significance that Pam's belief is correct. [2]

Jorge conducted a survey of 200 drivers. He asked two questions:

How long have you had your driving licence?
Do you wear a seat belt when driving?

The replies are summarized in the table below.

	Wear a seat belt	Do not wear a seat belt
Licence less than 2 years	38	42
Licence between 2 and 15 years	30	45
Licence more than 15 years	30	15

a. Jorge applies a χ^2 test at the 5% level to investigate whether wearing a seat belt is associated with the time a driver has had their licence. [8]

(i)   Write down the null hypothesis, H_0.

(ii) Write down the number of degrees of freedom.

(iii) Show that the expected number of drivers that wear a seat belt and have had their driving licence for more than 15 years is 22, correct to the nearest whole number.

(iv) Write down the χ^2 test statistic for this data.

(v)   Does Jorge accept H_0 ? Give a reason for your answer.

b. Consider the 200 drivers surveyed. One driver is chosen at random. Calculate the probability that [4]

(i)   this driver wears a seat belt;

(ii) the driver does not wear a seat belt, **given that** the driver has held a licence for more than 15 years.

c. Two drivers are chosen at random. Calculate the probability that [6]

(i)   both wear a seat belt.

(ii) at least one wears a seat belt.

The figure below shows the lengths in centimetres of fish found in the net of a small trawler.
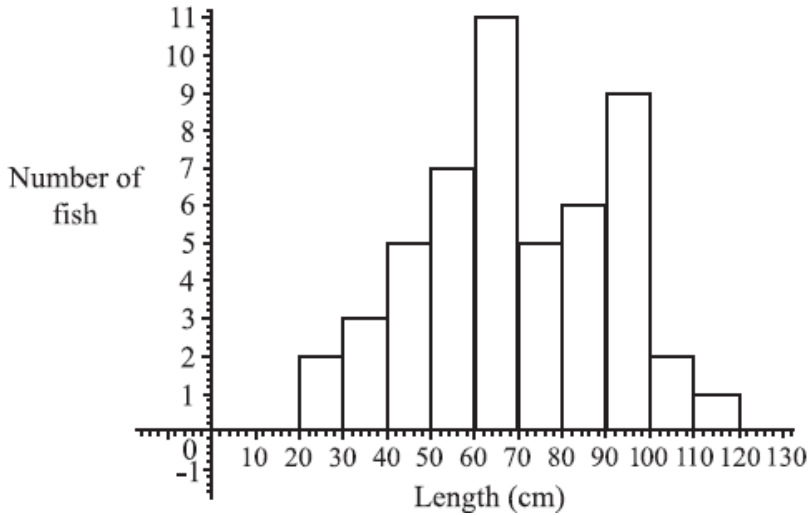
a. Find the total number of fish in the net. [2]

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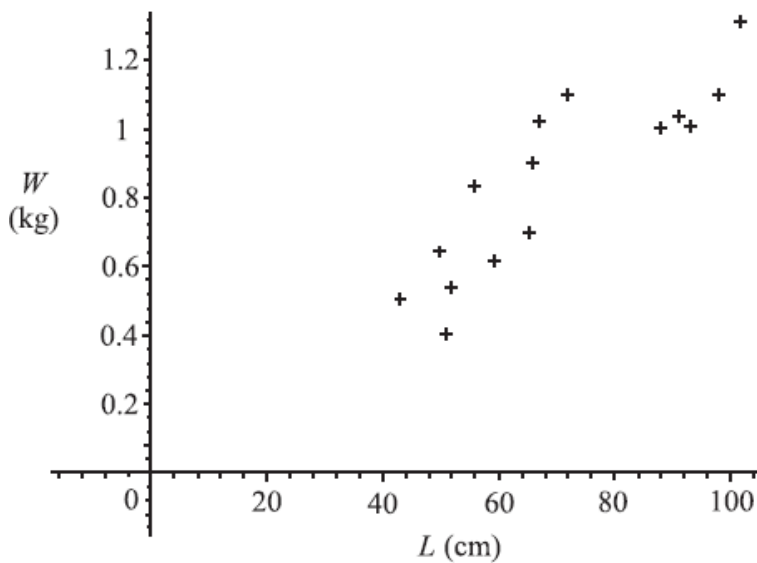
Exactly **two** of the following statements about the plot could be correct. Identify the two correct statements.

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 $W = 63.5L + 16.5$.

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			
		At most 500 km	Between 500 km and 5000 km	At least 5000 km	TOTAL
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)	$0 \le t < 15$	$15 \le t < 30$	$30 \le t < 45$	$45 \le t < 60$	$60 \le t < 75$	$75 \le t < 90$
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 [3]

mobile phones.

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 [2]

nearest whole number.

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.

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).

A survey of 400 people is carried out by a market research organization in two different cities, Buenos Aires and Montevideo. The people are asked which brand of cereal they prefer out of Chocos, Zucos or Fruti. The table below summarizes their responses.

	Chocos	Zucos	Fruti	Total
Buenos Aires	43	85	62	190
Montevideo	57	35	118	210
Total	100	120	180	400

The following table shows the cost in AUD of seven paperback books chosen at random, together with the number of pages in each book.

Book	1	2	3	4	5	6	7
Number of pages (x)	50	120	200	330	400	450	630
Cost (y AUD)	6.00	5.40	7.20	4.60	7.60	5.80	5.20

i.a. One person is chosen at random from those surveyed. Find the probability that this person [4]

    (i) does not prefer Zucos;

    (ii) prefers Chocos, given that they live in Montevideo.

i.b. Two people are chosen at random from those surveyed. Find the probability that they both prefer Fruti. [3]

i.c. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi- [1]

    squared test at the 5% level of significance is performed.

    State the null hypothesis.

i.d. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi- [1]

    squared test at the 5% level of significance is performed.

    State the number of degrees of freedom.

i.e. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi- [2]

    squared test at the 5% level of significance is performed.

    Show that the expected frequency for the number of people who live in Montevideo and prefer Zucos is 63.

i.f. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi- [2]

    squared test at the 5% level of significance is performed.

    Write down the chi-squared statistic for this data.

i.g. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi- [2]

    squared test at the 5% level of significance is performed.

    State whether the market research organization would accept the null hypothesis. Clearly justify your answer.

ii.a. Plot these pairs of values on a scatter diagram. Use a scale of $1\ \mathrm{cm}$ to represent 50 pages on the horizontal axis and $1\ \mathrm{cm}$ to represent $1\ \mathrm{AUD}$ [3]

    on the vertical axis.

ii.b. Write down the linear correlation coefficient, r, for the data. [2]

ii.c. Stephen wishes to buy a paperback book which has 350 pages in it. He plans to draw a line of best fit to determine the price. State whether or [2]

    not this is an appropriate method in this case and justify your answer.

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 $\mathrm{M}(\bar{S}, \bar{F})$ 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 [2]

line to estimate Monique's mark.

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

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]
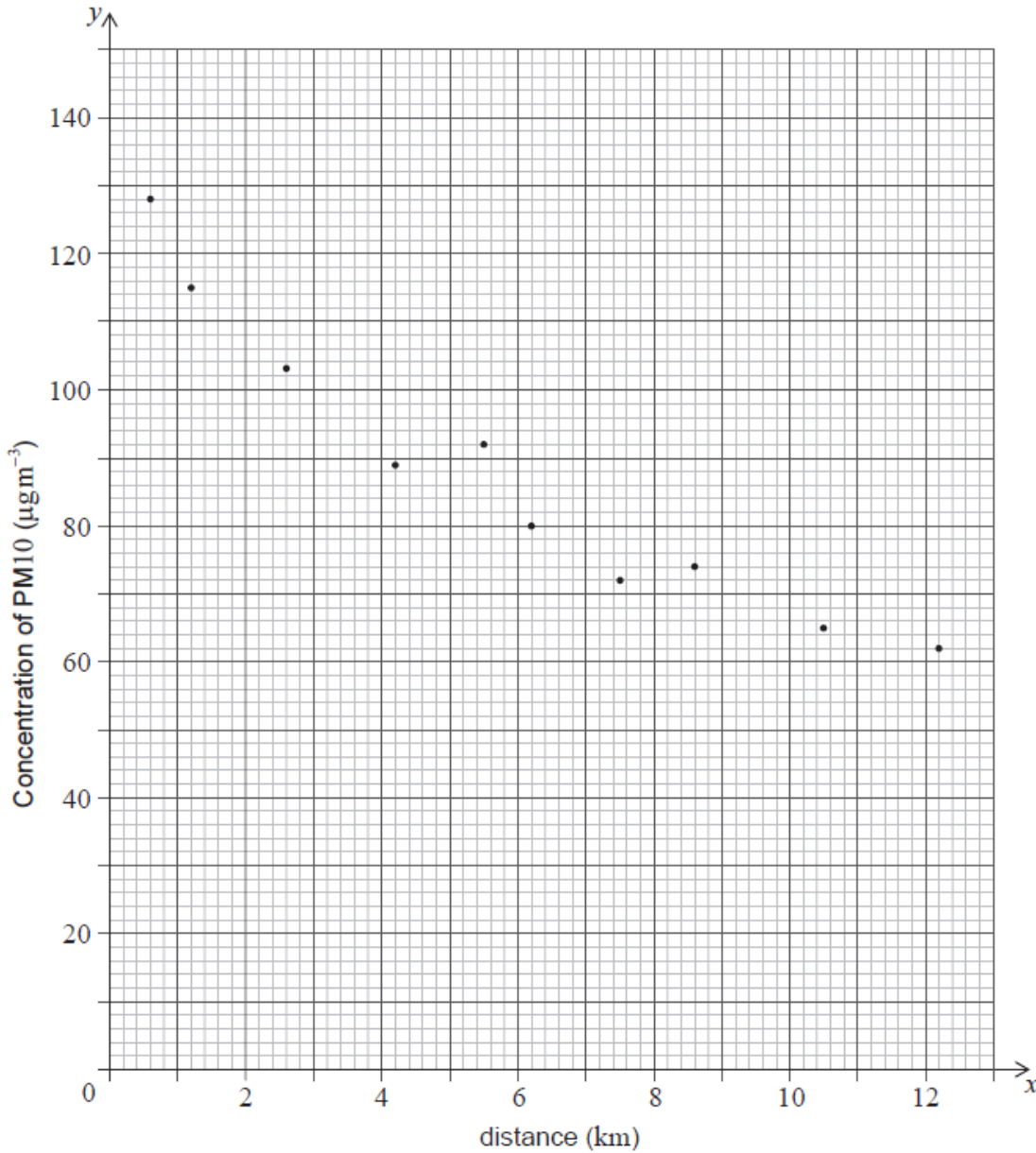
His data are also listed in the following table.

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 $14\,\mathrm{km}$ from the power plant. He uses the equation of the regression line to estimate the concentration of PM10 in   [4]

   the air at his school.

   i)   Calculate the value of Ernesto's estimate.

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

A manufacturer claims that fertilizer has an effect on the height of rice plants. He measures the height of fertilized and unfertilized plants. The

results are given in the following table.

Plant height	Fertilized plants	Unfertilized plants
>75 cm	115	80
50 – 75 cm	45	65
<50 cm	20	35

A chi-squared test is performed to decide if the manufacturer's claim is justified at the **1 %** level of significance.

The population of fleas on a dog after t days, is modelled by

$$N = 4 \times (2)^{\frac{t}{4}}, \ t \geqslant 0$$

Some values of N are shown in the table below.

t	0	4	8	12	16	20
N	p	8	16	32	q	128

i, a. Write down the null and alternative hypotheses for this test.   [2]

i, b. For the number of fertilized plants with height greater than 75 cm, show that the expected value is 97.5.   [3]

i, c. Write down the value of χ^2_{calc}.   [2]

i, d. Write down the number of degrees of freedom.   [1]

i, f. Is the manufacturer's claim justified? Give a reason for your answer.   [2]

ii, a. Write down the value of p.   [1]

ii, a. Write down the value of q.   [2]

ii, b. Using the values in the table above, draw the graph of N for $0 \leq t \leq 20$. Use 1 cm to represent 2 days on the horizontal axis and 1 cm to   [6]

   represent 10 fleas on the vertical axis.

ii, c. **Use your graph** to estimate the number of days for the population of fleas to reach 55.   [2]

The weight, W, of basketball players in a tournament is found to be normally distributed with a mean of 65 kg and a standard deviation of 5 kg.

The probability that a basketball player has a weight that is within 1.5 standard deviations of the mean is q.

A basketball coach observed 60 of her players to determine whether their performance and their weight were independent of each other. Her observations were recorded as shown in the table.

		Performance	
		Satisfactory	Excellent
Weight	Below average	6	10
	Average	7	15
	Above average	12	10

She decided to conduct a χ^2 test for independence at the 5% significance level.

a.i. Find the probability that a basketball player has a weight that is less than 61 kg. [2]

a.ii. In a training session there are 40 basketball players. [2]

   Find the expected number of players with a weight less than 61 kg in this training session.

b.i. Sketch a normal curve to represent this probability. [2]

b.ii. Find the value of q. [1]

c. Given that $P(W > k) = 0.225$, find the value of k. [2]

d.i. For this test state the null hypothesis. [1]

d.ii. For this test find the p-value. [2]

e. State a conclusion for this test. Justify your answer. [2]

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 [4]

   thermal units on the y-axis.

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 [2] 25% moisture content.

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 [2] on x to estimate the heat output in part (f).

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 [4] the y-axis.

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]

George leaves a cup of hot coffee to cool and measures its temperature every minute. His results are shown in the table below.

Time, t (minutes)	0	1	2	3	4	5	6
Temperature, y (°C)	94	54	34	24	k	16.5	15.25

a. Write down the decrease in the temperature of the coffee [3]

    (i) during the first minute (between $t = 0$ and $t = 1$) ;

    (ii) during the second minute;

    (iii) during the third minute.

b. Assuming the pattern in the answers to part (a) continues, show that $k = 19$. [2]

c. Use the **seven** results in the table to draw a graph that shows how the temperature of the coffee changes during the first six minutes. [4]

    Use a scale of 2 cm to represent 1 minute on the horizontal axis and 1 cm to represent 10 °C on the vertical axis.

d. The function that models the change in temperature of the coffee is $y = p\,(2^{-t}) + q$. [2]

    (i) Use the values $t = 0$ and $y = 94$ to form an equation in p and q.

    (ii) Use the values $t = 1$ and $y = 54$ to form a second equation in p and q.

e. Solve the equations found in part (d) to find the value of p and the value of q. [2]

f. The graph of this function has a horizontal asymptote. [2]

    Write down the equation of this asymptote.

g. George decides to model the change in temperature of the coffee with a linear function using correlation and linear regression. [4]

    Use the **seven** results in the table to write down

    (i) the correlation coefficient;

    (ii) the equation of the regression line y on t.

h. Use the equation of the regression line to estimate the temperature of the coffee at $t = 3$. [2]

i. Find the percentage error in this estimate of the temperature of the coffee at $t = 3$. [2]

A biologist is studying the relationship between the number of chirps of the Snowy Tree cricket and the air temperature. He records the chirp rate, x, of a cricket, and the corresponding air temperature, T, in degrees Celsius.

The following table gives the recorded values.

Cricket's chirp rate, x, (chirps per minute)	20	40	60	80	100	120
Temperature, T (°C)	8.0	12.8	15.0	18.2	20.0	21.1

a. Draw the scatter diagram for the above data. Use a scale of 2 cm for 20 chirps on the horizontal axis and 2 cm for 4°C on the vertical axis. [4]

b. Use your graphic display calculator to write down the Pearson's product–moment correlation coefficient, r, between x and T. [2]

c. Interpret the relationship between x and T using your value of r. [2]

d. Use your graphic display calculator to write down the equation of the regression line T on x. Give the equation in the form $T = ax + b$. [2]

e. Calculate the air temperature when the cricket's chirp rate is 70. [2]

f. Given that $\bar{x} = 70$, draw the regression line T on x on your scatter diagram. [2]

g. A forest ranger uses her own formula for estimating the air temperature. She counts the number of chirps in 15 seconds, z, multiplies this number by 0.45 and then she adds 10. [1]

Write down the formula that the forest ranger uses for estimating the temperature, T.
Give the equation in the form $T = mz + n$.

h. A cricket makes 20 chirps in **15** seconds. [6]

For this chirp rate

(i)   calculate an estimate for the temperature, T, **using the forest ranger's formula**;

(ii)   determine the actual temperature recorded by the biologist, **using the table above**;

(iii)   calculate the percentage error in the forest ranger's estimate for the temperature, compared to the actual temperature recorded by the biologist.