Bronze Questions

Calculator

The total mark for this section is 35

Q1

Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$$\sum x = 84.6$$

(a) Calculate the mean amount of rainfall during the whole 28 days.

(2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0
Keith corrects these figures.

(b) State, giving your reason, the effect this will have on the mean.

(2)

(Total for Question 1 is 4 marks)

Q2

Sara is investigating the variation in daily maximum gust, t kn, for Camborne in June and July 1987.

She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

(a) State the sampling technique Sara used.

(1)

(b) From your knowledge of the large data set, explain why this process may not generate a sample of size 20.

(1)

The data Sara collected are summarised as follows

$$n = 20 \qquad \sum t = 374 \qquad \sum t^2 = 7600$$

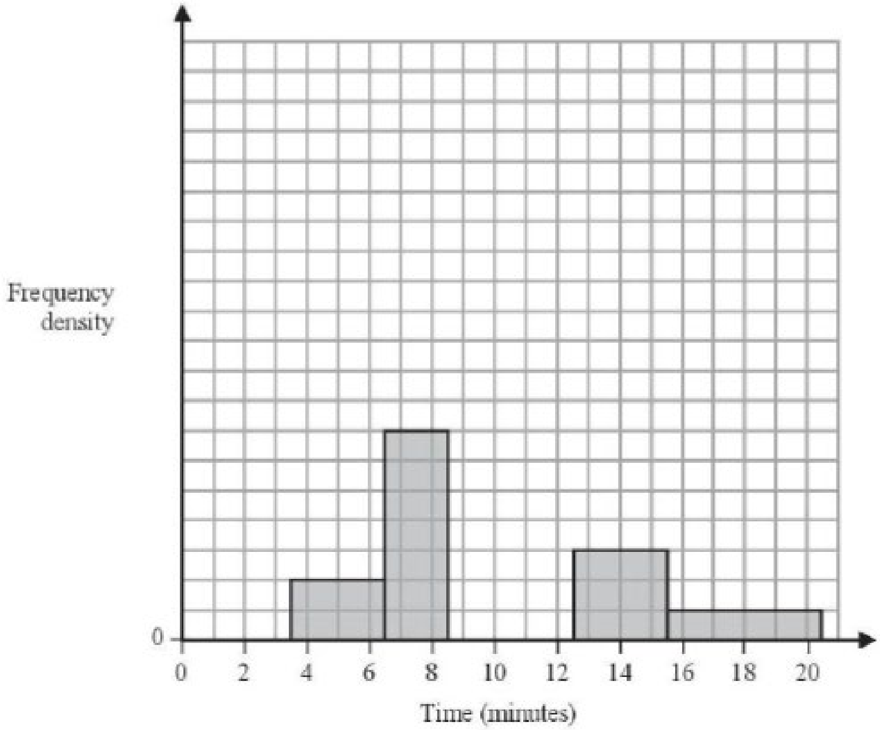
(c) Calculate the standard deviation.

(2)

(Total for Question 2 is 4 marks)

Q3

The partially completed histogram and the partially completed table show the time, to the nearest minute, that a random sample of motorists were delayed by roadworks on a stretch of motorway.



Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	17
10 – 12	45
13 – 15	9
16 – 20	

Estimate the percentage of these motorists who were delayed by the roadworks for between 8.5 and 13.5 minutes.

(Total for Question 3 is 5 marks)

Pearson

Q4

Helen is studying the daily mean wind speed for Camborne using the large data set from 1987.

The data for one month are summarised in Table 1 below.

Windspeed	n/a	6	7	8	9	11	12	13	14	16
Frequency	13	2	3	2	2	3	1	2	1	2

Table 1

(a) Calculate the mean for these data.

(1)

(b) Calculate the standard deviation for these data and state the units.

(2)

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

Month	*A*	*B*	*C*	*D*	*E*
Mean	7.58	8.26	8.57	8.57	11.57
Standard Deviation	2.93	3.89	3.46	3.87	4.64

Table 2

(c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

(2)

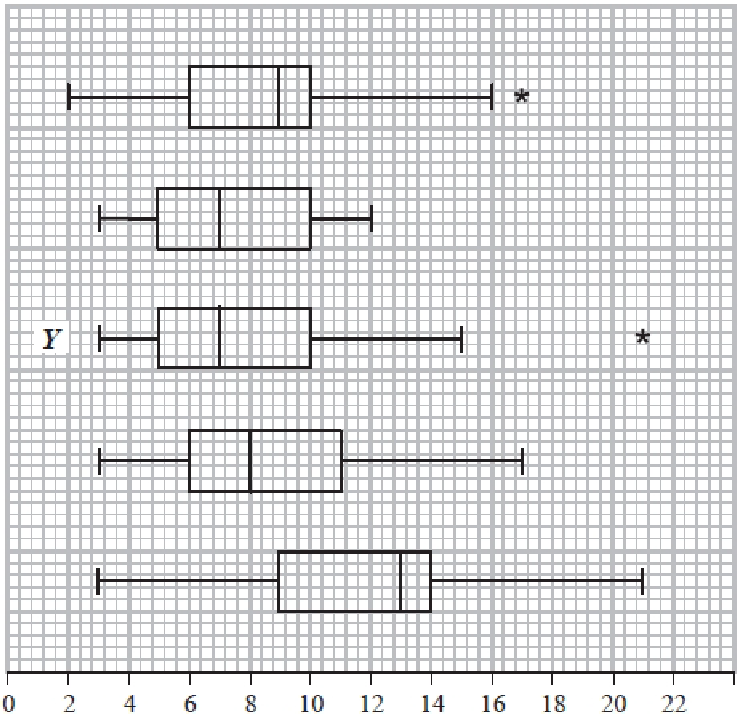
The data for these months are summarised in the box plots on the next page. They are not in month order or the same order as in Table 2.

(d) (i) State the meaning of the * symbol on some of the box plots.

(ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*.

(3)

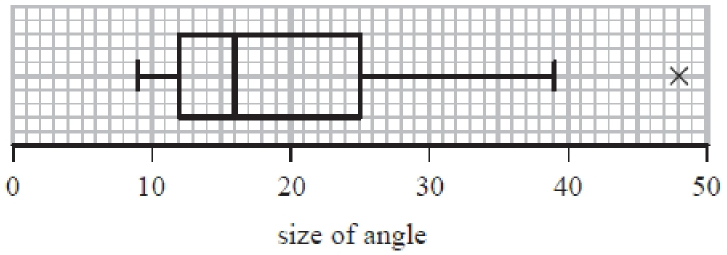
Pearson



(Total for Question 4 is 8 marks)

Q5

Each of 60 students was asked to draw a 20° angle without using a protractor. The size of each angle drawn was measured. The results are summarised in the box plot below.



(a) Find the range for these data.

(1)

(b) Find the interquartile range for these data.

(1)

The students were then asked to draw a 70° angle.
The results are summarised in the table below.

Angle, a, (degrees)	Number of students
$55 \leqslant a < 60$	6
$60 \leqslant a < 65$	15
$65 \leqslant a < 70$	13
$70 \leqslant a < 75$	11
$75 \leqslant a < 80$	8
$80 \leqslant a < 85$	7

(c) Use linear interpolation to estimate the size of the median angle drawn. Give your answer to 1 decimal place.

(2)

(d) Show that the lower quartile is 63°

(2)

For these data, the upper quartile is 75°, the minimum is 55° and the maximum is 84°

An outlier is an observation that falls either
more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

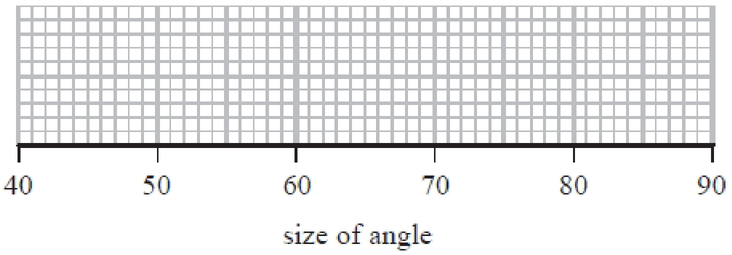
(e) (i) Show that there are no outliers for these data.

(ii) Draw a box plot for these data on the grid on the following page.

(5)

(f) State which angle the students were more accurate at drawing. Give reasons for your answer.

(3)



(Total for Question 5 is 14 marks)

Silver Questions

Calculator

The total mark for this section is 30

Q1

Sara was studying the relationship between rainfall, r mm, and humidity, h %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1 \quad Q_2 = 0.9 \quad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.

(a) Show that $r = 20.6$ is an outlier.

(1)

(b) Give a reason why Sara might

(i) include

(ii) exclude

this day's reading.

(2)

(Total for Question 1 is 3 marks)

Q2

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

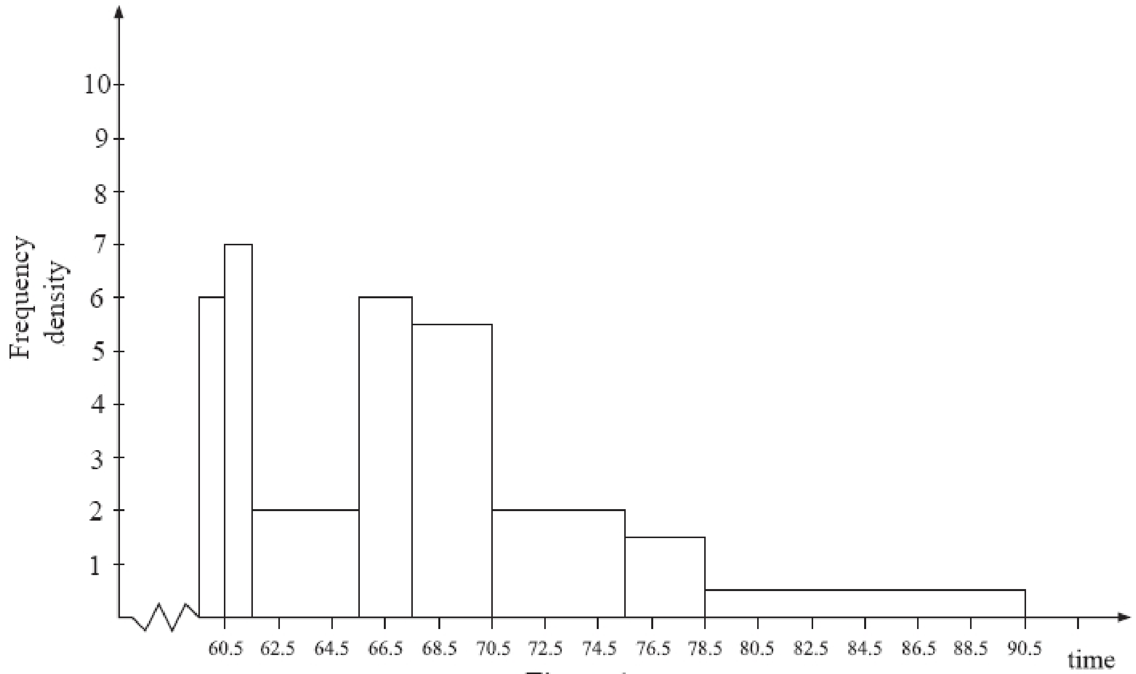


Figure 1

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

(Total for Question 2 is 5 marks)

Q3

Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	Frequency	Midpoint (x mm)
$0 \leqslant r < 0.5$	121	0.25
$0.5 \leqslant r < 1.0$	10	0.75
$1.0 \leqslant r < 5.0$	24	3.0
$5.0 \leqslant r < 10.0$	12	7.5
$10.0 \leqslant r < 30.0$	17	20.0

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

(Total for Question 3 is 8 marks)

Q4

The marks of a group of female students in a statistics test are summarised in Figure 1

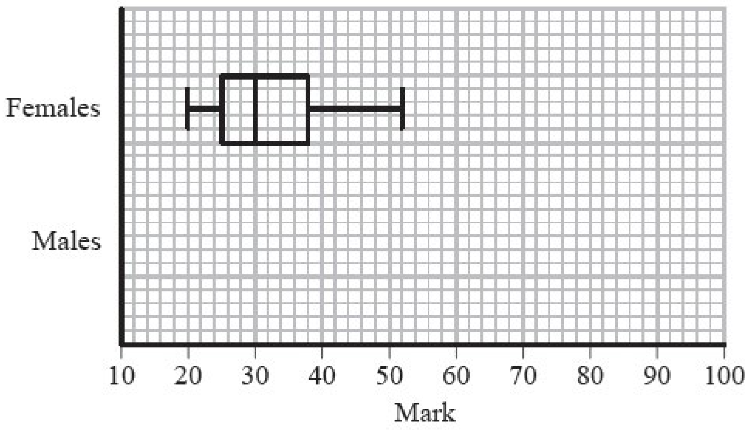


Figure 1

(a) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(b) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is either more than 1.5 × interquartile range above the upper quartile or more than 1.5 × interquartile range below the lower quartile.

(c) In the space provided on Figure 1 draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(d) Compare and contrast the marks of the male and the female students.

(2)

(Total for Question 4 is 11 marks)

Q5

The variable x was measured to the nearest whole number. Forty observations are given in the table below.

x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 - 15 class has a width of 2 cm and a height of 5 cm. For the 16 - 18 class find

(a) the width,

(1)

(b) the height of the bar representing this class.

(2)

(Total for Question 5 is 3 marks)

Gold Questions

Calculator

The total mark for this section is 33

Q1

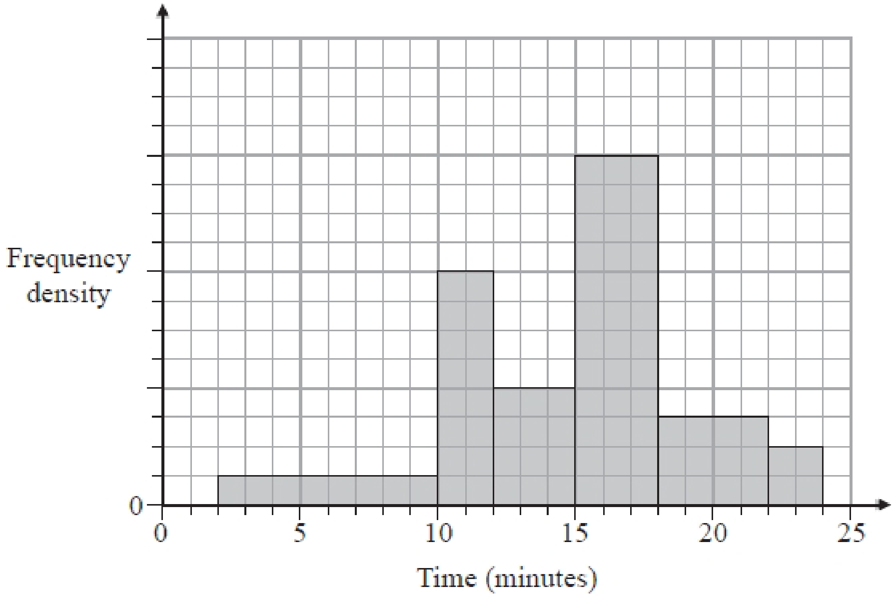


Figure 1

The histogram in Figure 1 shows the times taken to complete a crossword by a random sample of students.

The number of students who completed the crossword in more than 15 minutes is 78.

Estimate the percentage of students who took less than 11 minutes to complete the crossword.

(Total for Question 1 is 4 marks)

Q2

The mark, x, scored by each student who sat a statistics examination is coded using

$$y = 1.4x - 20$$

The coded marks have mean 60.8 and standard deviation 6.60

Find the mean and the standard deviation of x.

(4)

(Total for Question 2 is 4 marks)

Q3

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data.

(5)

(b) Use linear interpolation to estimate the value of the median.

(2)

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures.

(1)

(d) Estimate the interquartile range of this distribution.

(2)

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data.

(1)

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in parts (a), (b), (c) and (d).

(3)

(Total for Question 3 is 14 marks)

Q4

A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	**Frequency (f)**	**Weight midpoint (x)**
$0 \leqslant w < 2$	1	1
$2 \leqslant w < 3$	8	2.5
$3 \leqslant w < 3.5$	17	3.25
$3.5 \leqslant w < 4$	17	3.75
$4 \leqslant w < 5$	7	4.5

[You may use $\sum fx^2 = 611.375$]

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leq w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leq w < 3.5$

(3)

(b) Use linear interpolation to estimate the median weight of these babies.

(2)

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies.

(3)

A newborn baby weighing 3.43 kg is born at the hospital.

(f) Without carrying out any further calculations, state, giving a reason, what effect the addition of this newborn baby to the sample would have on your estimate of the

(i) mean,
(ii) standard deviation.

(3)

(Total for Question 4 is 11 marks)

<u>Bronze Questions</u>

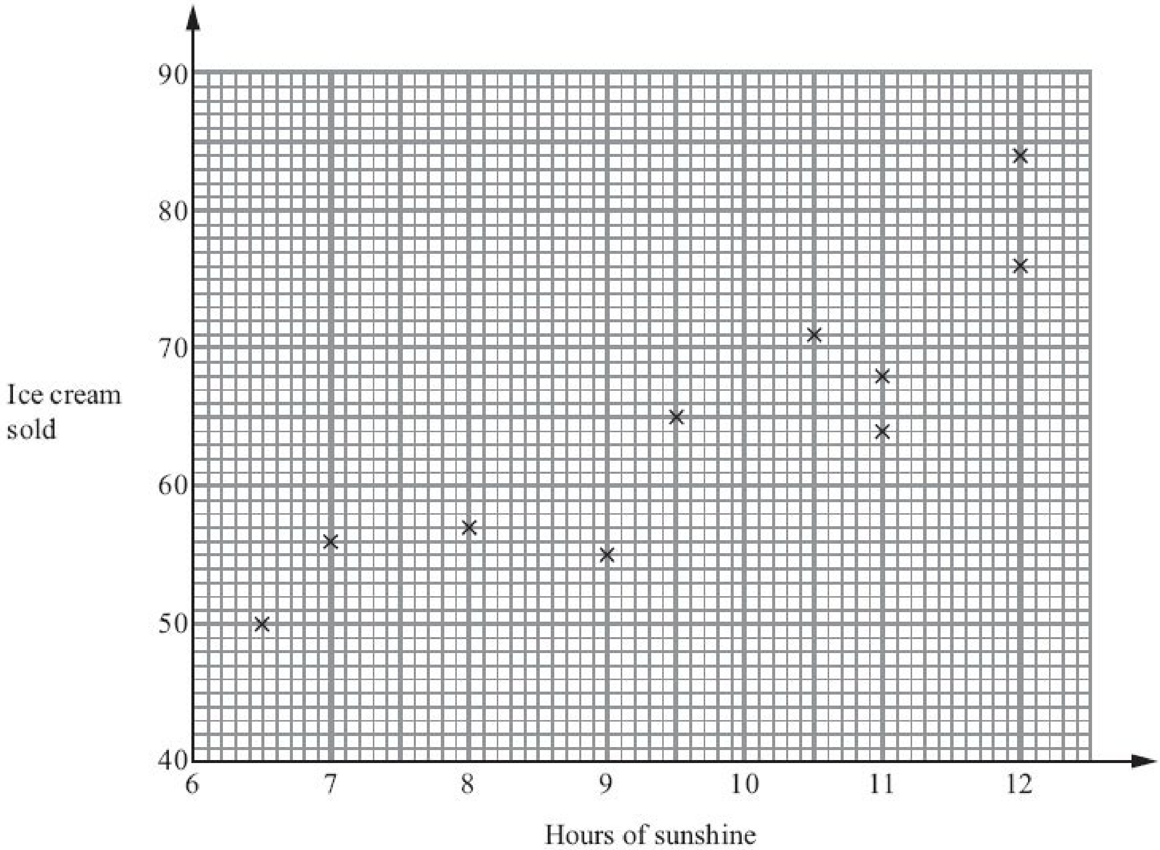
Calculator

The total mark for this section is 28

Q1

A beach cafe sells ice creams.
Each day the manager records the number of hours of sunshine and the number of ice creams sold.

The scatter graph shows this information.



On another day there were 11.5 hours of sunshine and 73 ice creams sold.

(a) Show this information on the scatter graph.

(1)

(b) Describe the relationship between the number of hours of sunshine and the number of ice creams sold.

(1)

One day had 10 hours of sunshine.

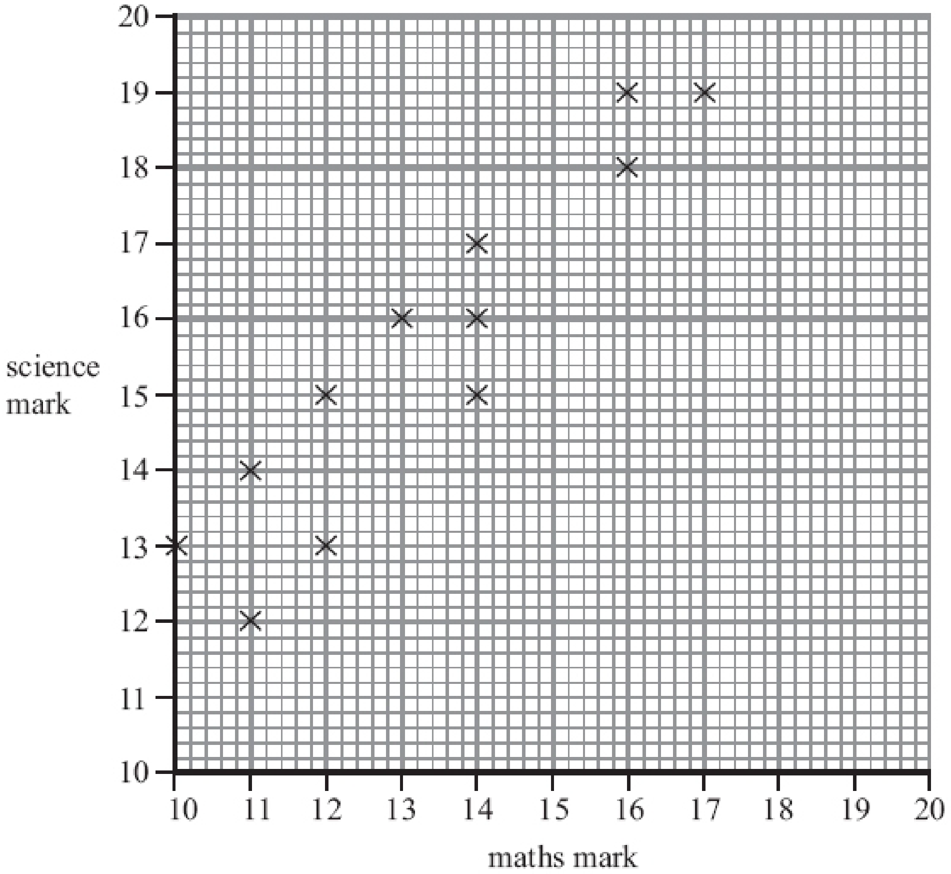
(c) Estimate how many ice creams were sold.

(2)

(Total for Question 1 is 4 marks)

Q2

Mr Kent's students did a maths test and a science test.
The scatter graph shows the marks of 12 of these students.



The table shows the marks of two more students.

Name	**maths**	**science**
Masood	12	14
Nimer	17	20

(a) Show this information on the scatter graph.

(1)

(b) What type of correlation does this scatter graph show?

.

(1)

David did the maths test.
He was absent for the science test.

David's mark in the maths test was 15

(c) Estimate a science mark for David.

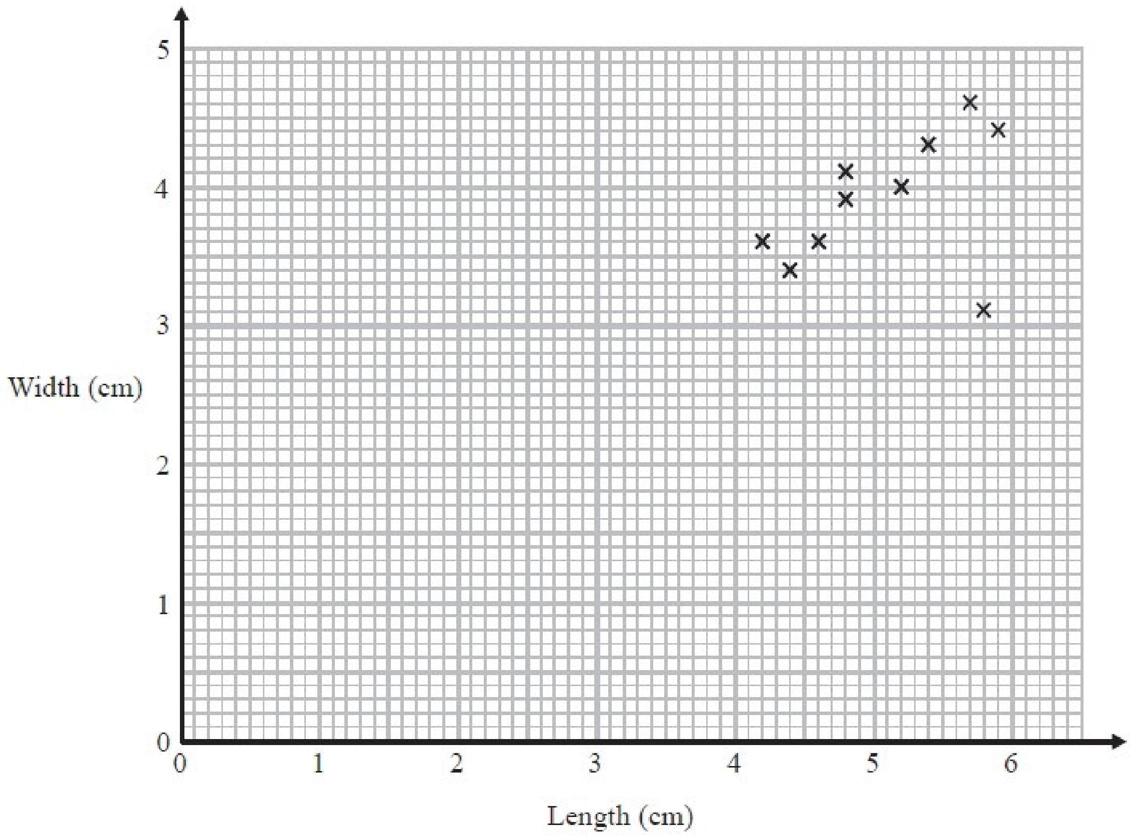
(2)

(Total for Question 2 is 4 marks)

Q3

Katie measured the length and the width of each of 10 pine cones from the same tree.

She used her results to draw this scatter graph.



(a) Describe one improvement Katie can make to her scatter graph.

(1)

The point representing the results for one of the pine cones is an outlier.

(b) Explain how the results for this pine cone differ from the results for the other pine cones.

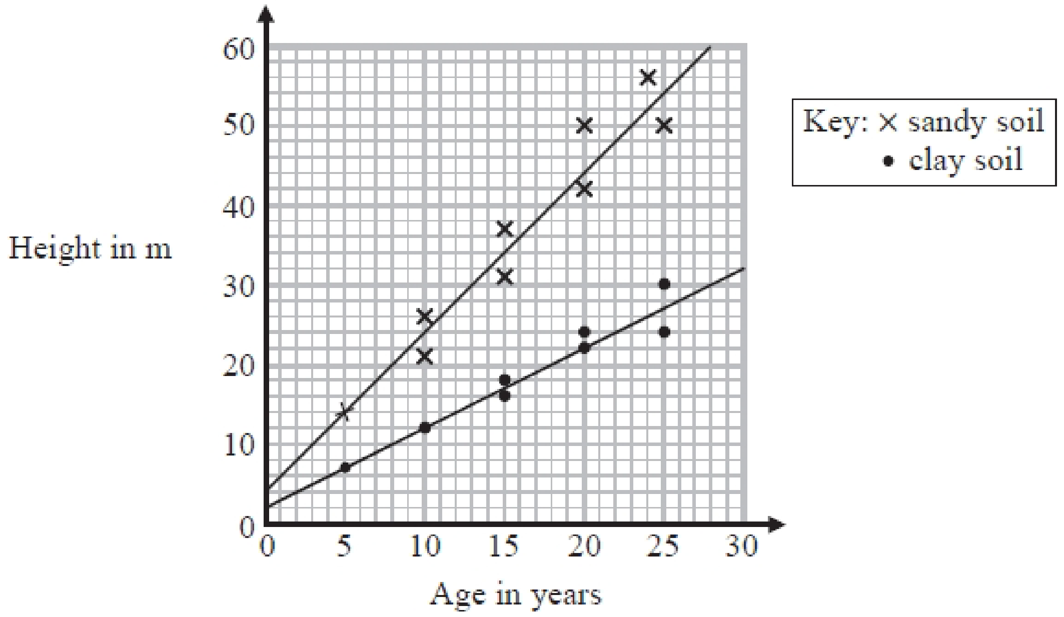
(1)

(Total for Question 3 is 2 marks)

Q4

Bill wants to compare the heights of pine trees growing in sandy soil with the heights of pine trees growing in clay soil.

The scatter diagram gives some information about the heights and the ages of some pine trees.



(a) Describe the relationship between the height of pine trees and the age of pine trees growing in sandy soil.

(1)

A pine tree growing in clay soil is 18 years old.

(b) Find an estimate for the height of this tree.

(1)

A pine tree is growing in sandy soil.

(c) Work out an estimate for how much the height of this tree increases in a year.

(2)

(d) Compare the rate of increase of the height of trees growing in clay soil with the rate of increase of the height of trees growing in sandy soil.

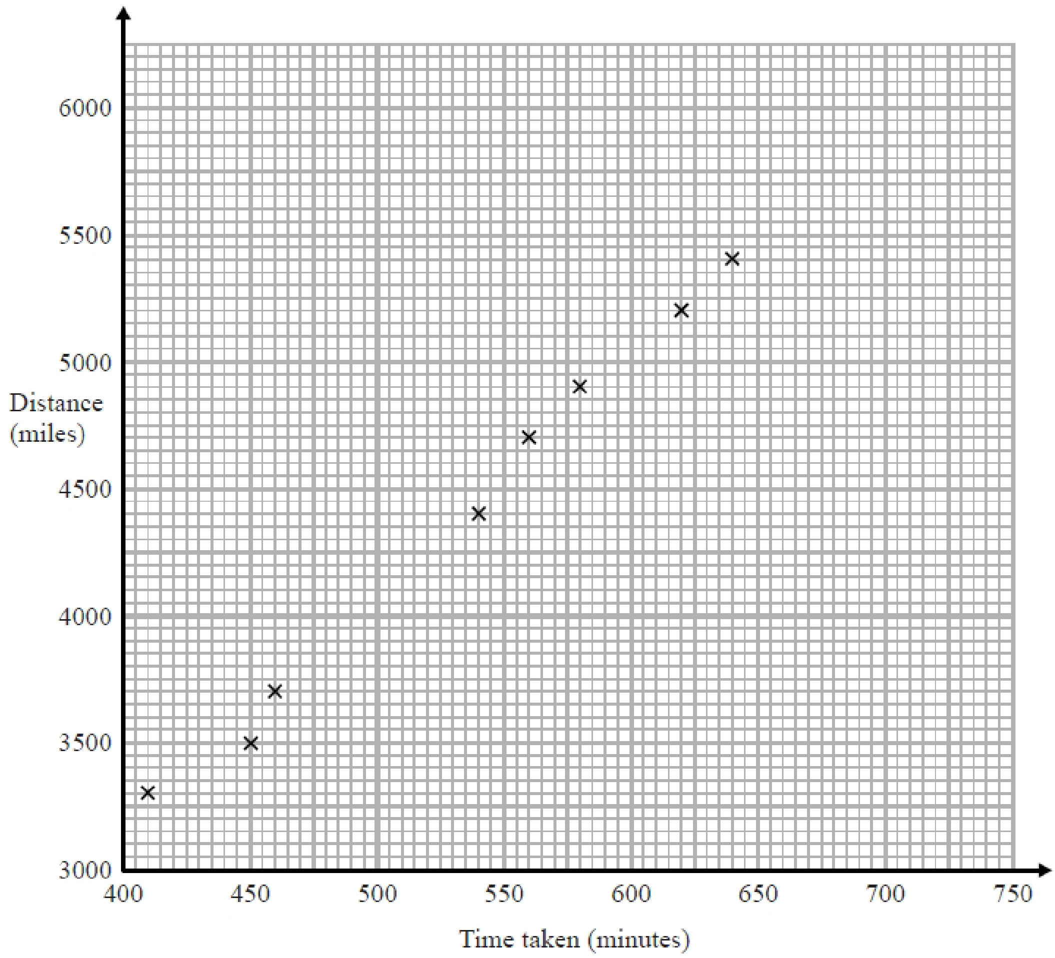
(2)

(Total for Question 4 is 6 marks)

Q5

Oliver records the distance from London to each of eight cities in the USA.
He also records the time taken to fly from London to each of these cities.

The scatter graph shows this information.



Chicago is a city in the USA.
Chicago is 4000 miles from London.

(a) (i) By drawing a line of best fit, find an estimate for the time taken to fly from London to Chicago.

(2)

(ii) Why is your answer to part (i) only an estimate?

(1)

(b) (i) Calculate the gradient of your line of best fit.

(2)

(ii) Give an interpretation of the gradient of your line of best fit.

(1)

(Total for Question 5 is 6 marks)

Pearson

Q6

The age, t years, and weight, w grams, of each of 10 coins were recorded.

Given that the equation of the regression line of w on t is $w = 11.6 - 0.0263t$

(a) State, with a reason, which variable is the explanatory variable.

(2)

(b) Using this model, estimate

(i) the weight of a coin which is 5 years old,
(ii) the effect of an increase of 4 years in age on the weight of a coin.

(2)

It was discovered that a coin in the original sample, which was 5 years old and weighed 20 grams, was a fake.

(c) State, without any further calculations, whether the exclusion of this coin would increase or decrease the value of the
correlation. Give a reason for your answer.

(2)

(Total for Question 6 is 6 marks)

Silver Questions

Calculator

The total mark for this section is 28

Q1

A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

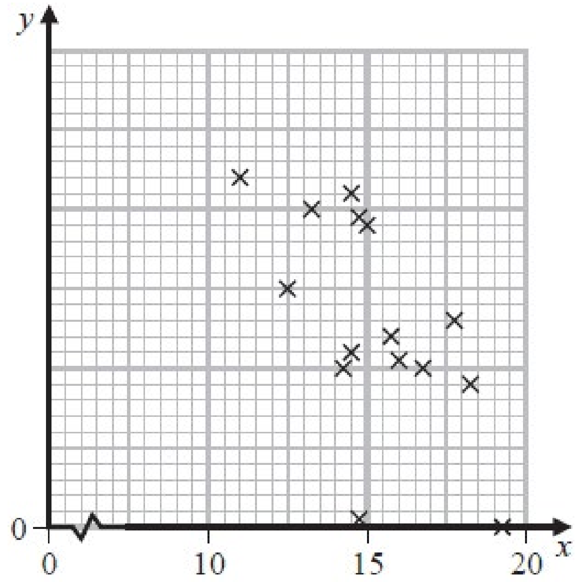


Figure 1

(a) Describe the correlation.

(1)

The variable on the x-axis is Daily Mean Temperature measured in °C.

(b) Using your knowledge of the large data set,

(i) suggest which variable is on the y-axis,

(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a negative correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(c) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

(Total for Question 1 is 4 marks)

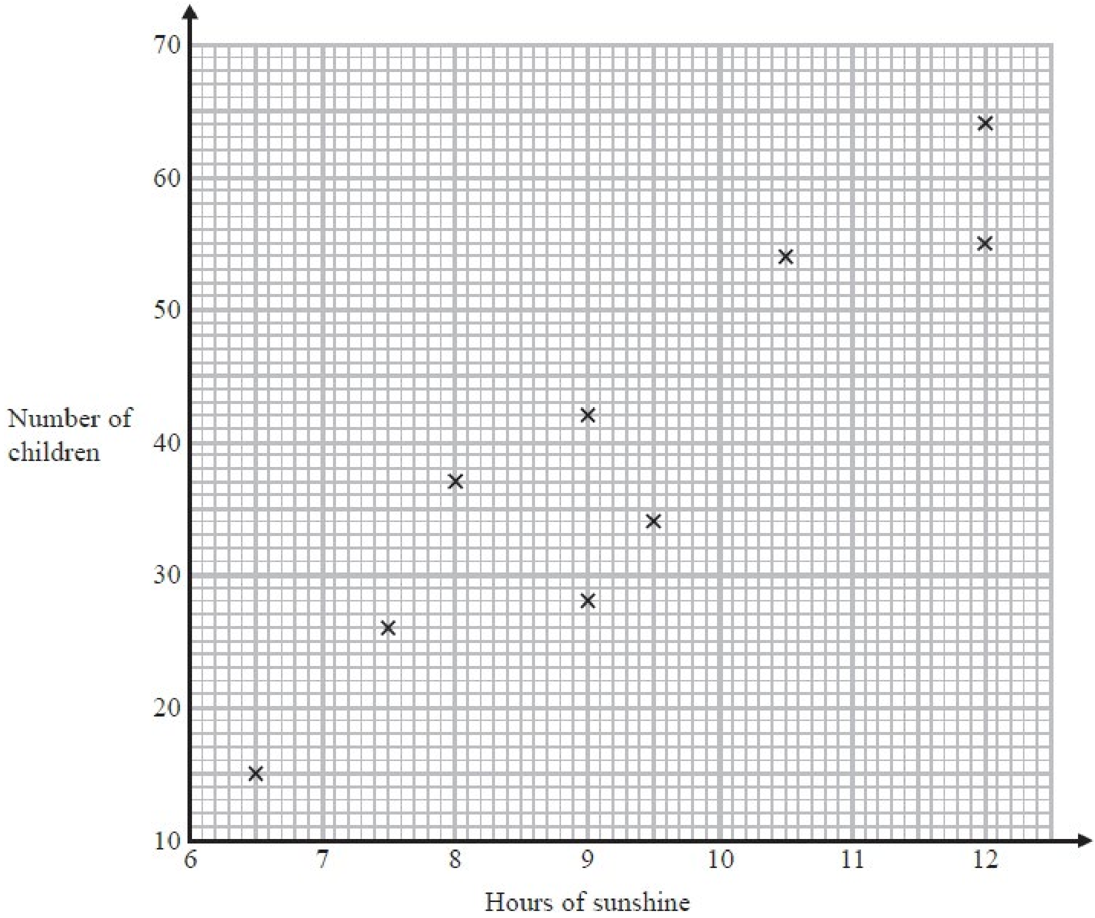
Pearson

Q2

Sally looks after a children's paddling pool in a park.

Each day, Sally records the number of hours of sunshine and the number of children who use the paddling pool.

The scatter graph shows this information.



(a) Describe the correlation between the number of children who use the paddling pool and the number of hours of sunshine.

(1)

One day there were 10 hours of sunshine.

(b) Estimate how many children used the paddling pool.

(2)

Pearson

On another day, there were 6.5 hours of sunshine and 45 children used the pool.

(c) (i) Show this information on the scatter graph.

This point is isolated on the scatter graph.

(ii) Explain what may have happened on this day.

(2)

(Total for Question 2 is 5 marks)

Q3

A scientist is researching whether or not birds of prey exposed to pollutants lay eggs with thinner shells. He collects a random sample of egg shells from each of 6 different nests and tests for pollutant level, p, and measures the thinning of the shell, t. The results are shown in the table below.

p	3	8	30	25	15	12
t	1	3	9	10	5	6

(a) Draw a scatter diagram to represent these data.

(2)

(b) Explain why a linear regression model may be appropriate to describe the relationship between p and t.

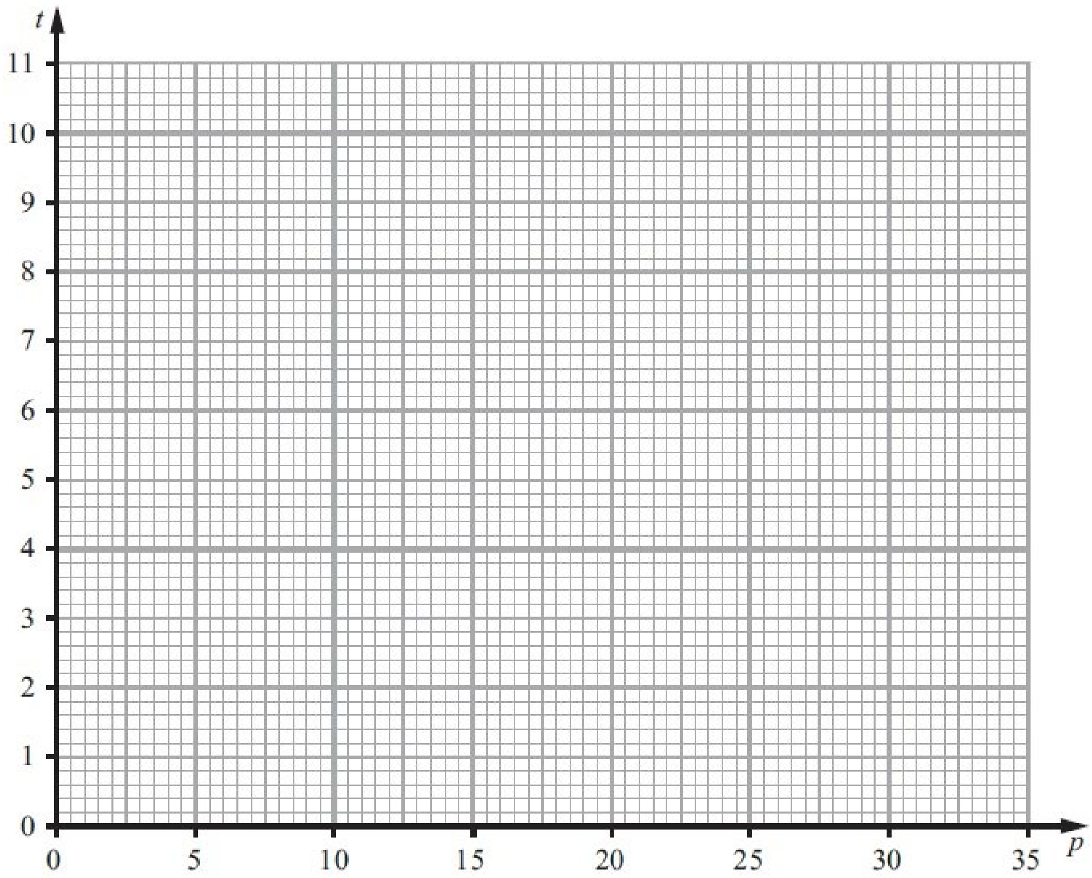
(1)

The scientist reviews similar studies and finds that pollutant levels above 16 are likely to result in the death of a chick soon after hatching.

Given that $t = 0.741+0.318p$

(c) Estimate the minimum thinning of the shell that is likely to result in the death of a chick.

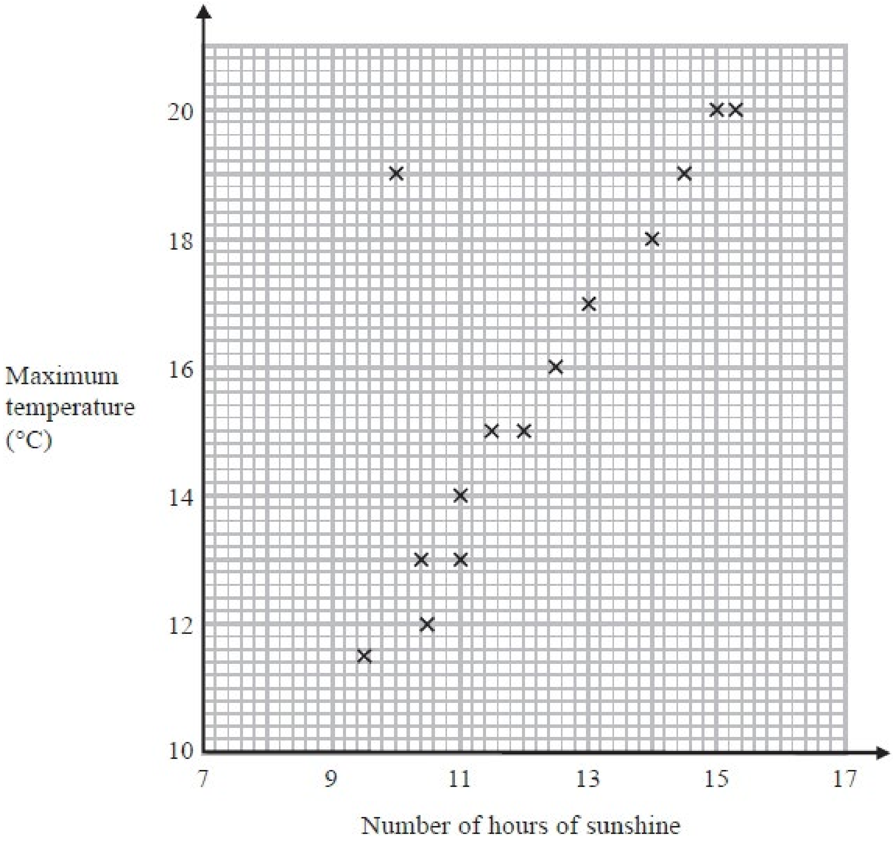
(2)



(Total for Question 3 is 5 marks)

Q4

The scatter graph shows the maximum temperature and the number of hours of sunshine in fourteen British towns on one day.



One of the points is an outlier.

(a) Write down the coordinates of this point.

(1)

(b) For all the other points write down the type of correlation.

(1)

On the same day, in another British town, the maximum temperature was 16.4°C.

(c) Estimate the number of hours of sunshine in this town on this day.

(2)

A weatherman says,

"Temperatures are higher on days when there is more sunshine."

(d) Does the scatter graph support what the weatherman says?

Give a reason for your answer.

(1)

(Total for Question 4 is 5 marks)

Q5

Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £ w, and the average weekly temperature, t °C, for 8 weeks during the summer.
There is a negative correlation for these data.

(a) Suggest a possible reason for this correlation.

(1)

Tessa suggests that a linear regression model could be used to model these data.

(b) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion.

(1)

(c) State, giving a reason, which variable would be the explanatory variable.

(1)

Tessa calculated the linear regression equation as $w = 10\,755 - 171t$

(d) Give an interpretation of the gradient of this regression equation.

(1)

(Total for Question 5 is 4 marks)

Q6

Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

Date	Daily Maximum Relative Humidity	Daily Mean Visibility
Units	**%**	
10/06/1987	90	5300
28/06/1987	100	0

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

(a) Write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility.

(1)

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

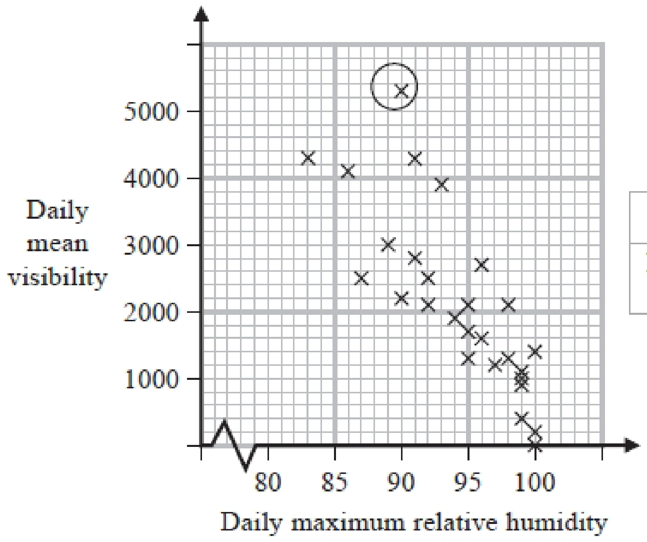


Figure 2

	Q_1	IQR
Daily mean visibility	1100	1600
Daily maximum relative humidity (%)	92	8

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above Q_3 or more than 1.5 times the interquartile range below Q_1.

(b) Show that the point circled on the scatter diagram is an outlier for visibility.

(2)

(c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity.

(1)

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the *x*–axis.

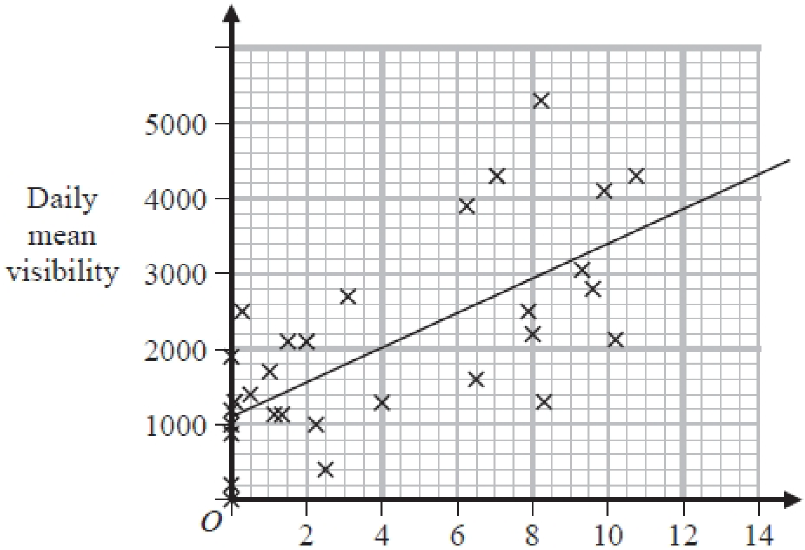


Figure 3

(d) Using your knowledge of the large data set, suggest which variable the *x*-axis on this scatter diagram represents.

(1)

(Total for Question 6 is 6 marks)

Gold Questions

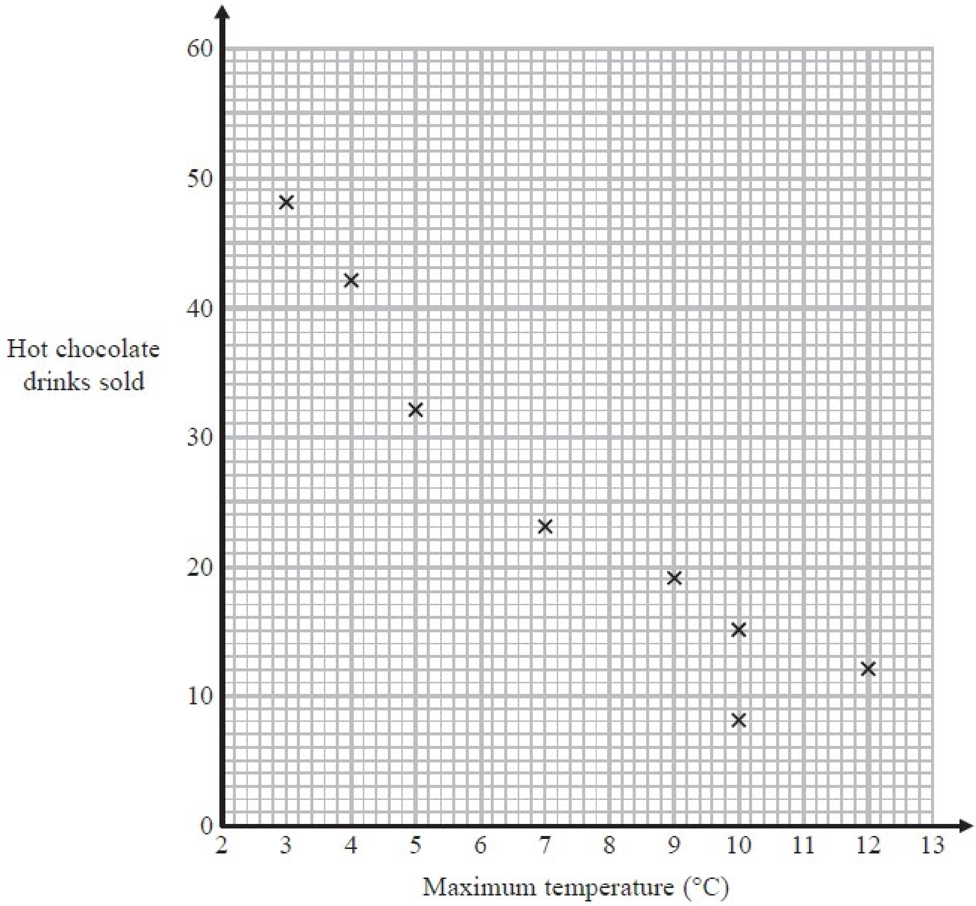
Calculator

The total mark for this section is 26

Q1

Carlos has a cafe in Clacton.
Each day, he records the maximum temperature in degrees Celsius (°C) in Clacton and the number of hot chocolate drinks sold.

The scatter graph shows this information.



On another day the maximum temperature was 6 °C and 35 hot chocolate drinks were sold.

(a) Show this information on the scatter graph.

(1)

(b) Describe the relationship between the maximum temperature and the number of hot chocolate drinks sold.

(1)

(c) Draw a line of best fit on the scatter diagram.

(1)

One day the maximum temperature was 8 °C.

(d) Use your line of best fit to estimate how many hot chocolate drinks were sold.

(1)

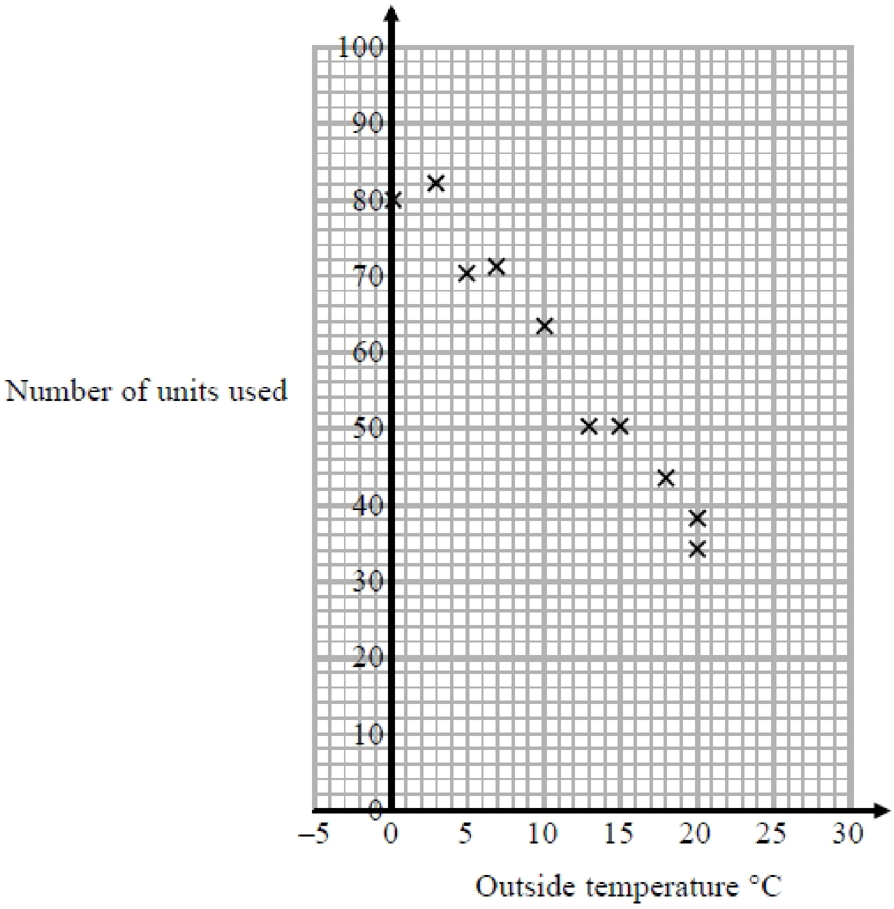
(Total for Question 1 is 4 marks)

Pearson

Q2

In a survey, the outside temperature and the number of units of electricity used for heating were recorded for ten homes.

The scatter diagram shows this information.



Molly says,

> "On average the number of units of electricity used for heating decreases by 4 units for each °C increase in outside temperature."

(a) Is Molly right?
Show how you get your answer.

(3)

(b) You should **not** use a line of best fit to predict the number of units of electricity used for heating when the outside temperature is 30 °C.

Give one reason why.

(1)

(Total for Question 2 is 4 marks)

Pearson

Q3

A meteorologist believes that there is a positive correlation between the daily mean windspeed, w kn, and the daily mean temperature, t °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below.

t	13.3	16.2	15.7	16.6	16.3	16.4	19.3	17.1	13.2
w	7	11	8	11	13	8	15	10	11

(a) Explain why a linear regression model based on these data is unreliable on a day when the mean temperature is 24 °C

(1)

Using the same 9 days a location from the large data set gave $\bar{t} = 27.2$ and $\bar{w} = 3.5$

(b) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics.

(1)

(Total for Question 3 is 2 marks)

Q4

A large company is analysing how much money it spends on paper in its offices every year. The number of employees, x, and the amount of money spent on paper, p (£ hundreds), in 8 randomly selected offices are given in the table below.

x	8	9	12	14	7	3	16	19
p (£ hundreds)	40.5	36.1	30.4	39.4	32.6	31.1	43.4	45.7

(a) Given that $p = 28.3 + 0.824x$

Estimate the amount of money spent on paper in an office with 10 employees.

(2)

(b) Explain the effect each additional employee has on the amount of money spent on paper.

(1)

Later the company realised it had made a mistake in adding up its costs, p. The true costs were actually half of the values recorded. The product moment correlation coefficient and the equation of the linear regression line are recalculated using this information.

(c) Write down the new value of the gradient of the regression line

(1)

(Total for Question 4 is 4 marks)

Q5

A company is introducing a job evaluation scheme. Points (x) will be awarded to each job based on the qualifications and skills needed and the level of responsibility. Pay (£y) will then be allocated to each job according to the number of points awarded.

Before the scheme is introduced, a random sample of 8 employees was taken and the linear regression equation of pay on points was $y = 4.5x - 47$

(a) Describe the correlation between points and pay.

(1)

(b) Give an interpretation of the gradient of this regression line.

(1)

(c) Explain why this model might not be appropriate for all jobs in the company.

(1)

(Total for Question 5 is 3 marks)

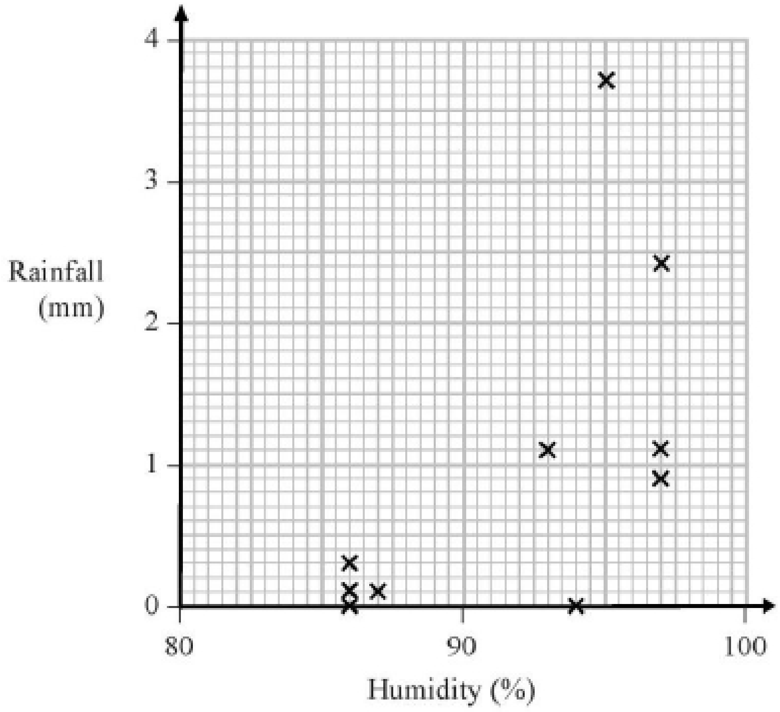
Q6

Sara was studying the relationship between rainfall, r mm, and humidity, h %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of r and h.



(a) Give an interpretation of the correlation between rainfall and humidity.

(1)

The equation of the regression line of r on h for these 10 days is $r = -12.8 + 0.15h$

(b) Give an interpretation of the gradient of this regression line.

(1)

(c) (i) Comment on the suitability of Sara's sampling method for this study.

(ii) Suggest how Sara could make better use of the large data set for her study.

(2)

(Total for Question 6 is 4 marks)

Q7

A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

(a) Describe how this sample could be taken.

(3)

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s$$

(b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

(1)

(c) Describe one limitation of this regression model.

(1)

(Total for Question 7 is 5 marks)

Bronze Questions

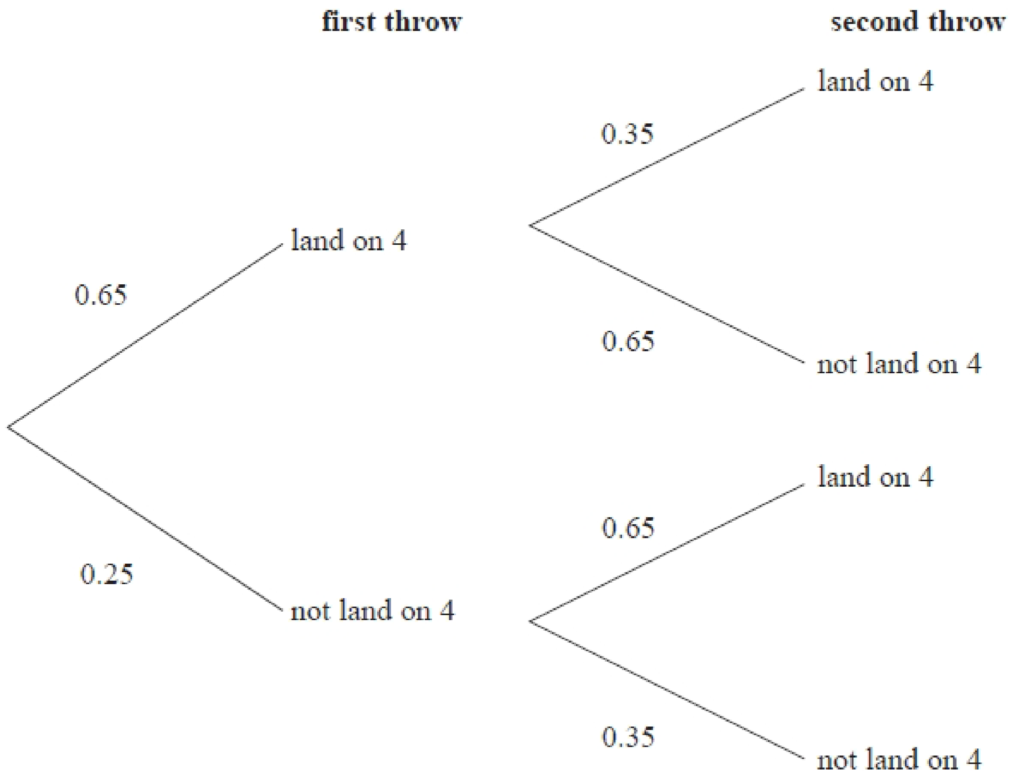
Calculator

The total mark for this section is 29

Q1

When a biased 6-sided dice is thrown once, the probability that it will land on 4 is 0.65
The biased dice is thrown twice.

Amir draws this probability tree diagram.
The diagram is **not** correct.



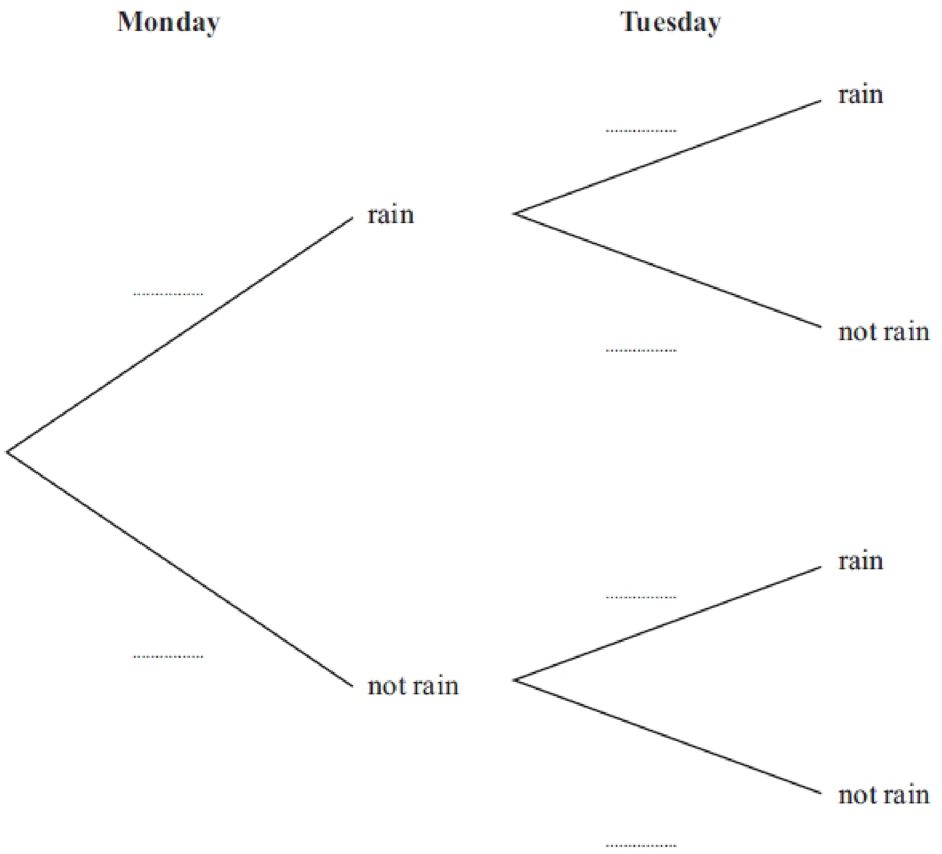
Write down **two** things that are wrong with the probability tree diagram.

(Total for Question 1 is 2 marks)

Q2

The probability that it will rain on Monday is 0.6

When it rains on Monday, the probability that it will rain on Tuesday is 0.8
When it does **not** rain on Monday, the probability that it will rain on Tuesday is 0.5



(a) Complete the probability tree diagram.

(2)

(b) Work out the probability that it will rain on both Monday and Tuesday.

.

(2)

(c) Work out the probability that it will rain on at least one of the two days.

(3)

(Total for Question 2 is 7 marks)

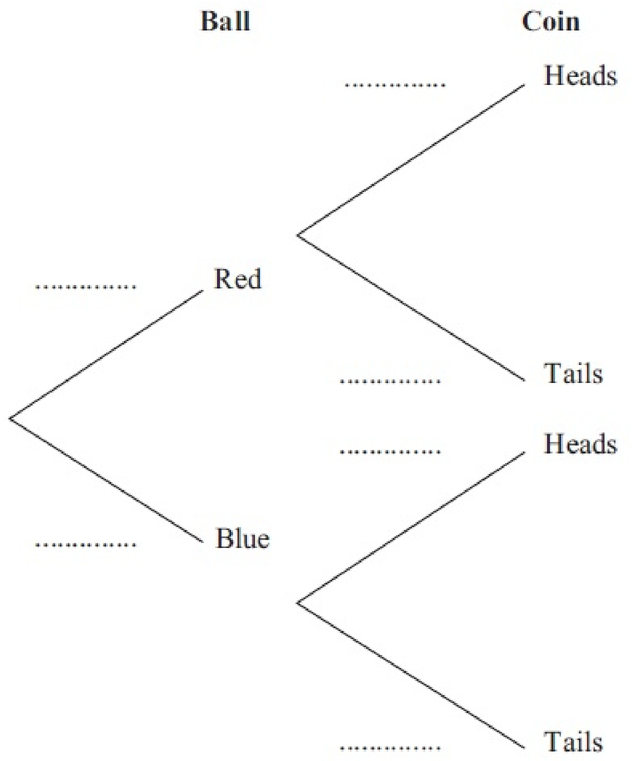
Q3

An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.

When a blue ball is selected a fair coin is spun.

(a) Complete the tree diagram below to show the possible outcomes and associated probabilities.



(2)

Shivani selects a ball and spins the appropriate coin.

(b) Find the probability that she obtains a head.

(2)

Shivani and Tom each repeat this experiment.

(c) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects.

(3)

(Total for Question 3 is 7 marks)

Pearson

Q4

Sami asked 50 people which drinks they liked from tea, coffee and milk.

All 50 people like at least one of the drinks

19 people like all three drinks.

16 people like tea and coffee but do **not** like milk.

21 people like coffee and milk.

24 people like tea and milk.

40 people like coffee.

1 person likes only milk.

Sami selects at random one of the 50 people.

(a) Work out the probability that this person likes tea.

(4)

(b) Given that the person selected at random from the 50 people likes tea, find the probability that this person also likes exactly one other drink.

(2)

(Total for Question 4 is 6 marks)

Q5

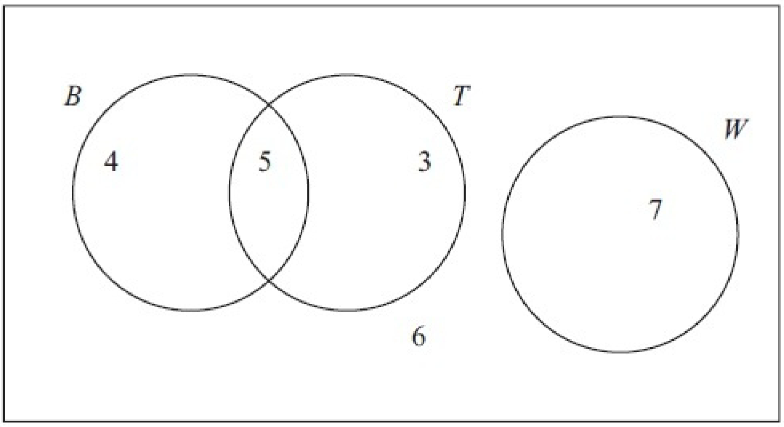


Figure 1

Figure 1 shows how 25 people travelled to work.

Their travel to work is represented by the events

B bicycle
T train
W walk

(a) Write down 2 of these events that are mutually exclusive. Give a reason for your answer.

(2)

(b) Determine whether or not *B* and *T* are independent events.

(3)

One person is chosen at random.

Find the probability that this person

(c) Walks to work,

(1)

(d) Travels to work by bicycle and train.

(1)

(Total for Question 5 is 7 marks)

Silver Questions

Calculator

The total mark for this section is 31

Q1

A jar contains 2 red, 1 blue and 1 green bead. Two beads are drawn at random from the jar without replacement.

(a) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities. State your probabilities clearly.

(3)

(b) Find the probability that a blue bead and a green bead are drawn from the jar.

(2)

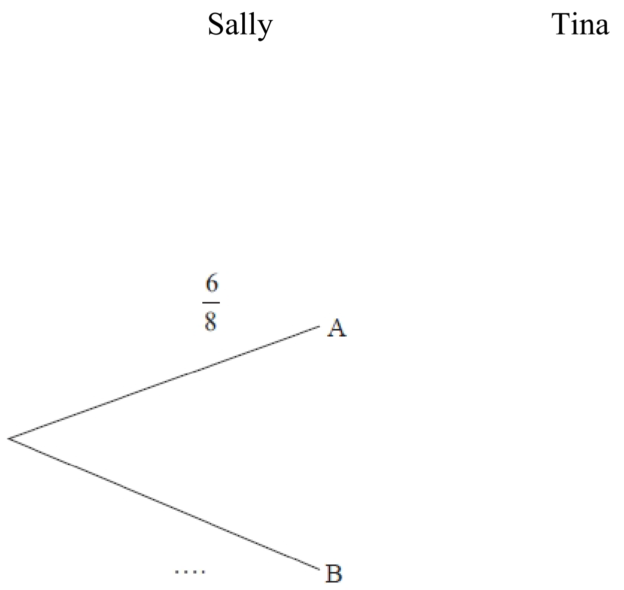
(Total for Question 1 is 5 marks)

Q2

There are 8 counters in a box.
The letter A is on 6 of the counters.
The letter B is on the other 2 counters.

Sally takes at random a counter from the box.
She keeps the counter.
Then Tina takes at random a counter from the box.

(a) Complete the probability tree diagram.



(3)

(b) Work out the probability that both Sally and Tina take a counter with the letter A on it.

(2)

(c) Work out the probability that at least one counter with the letter A on it is taken.

(3)

(Total for Question 2 is 8 marks)

Pearson

Q3

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

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

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

(6)

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

(b) None of the three wines,

(1)

(c) Wine A but not wine B,

(2)

(d) Any wine in the survey except wine C,

(2)

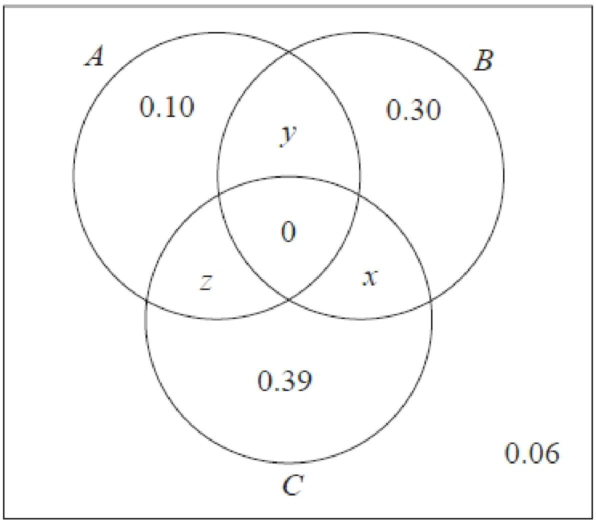
(e) Exactly two of the three kinds of wine.

(2)

(Total for Question 3 is 3 marks)

Q4

The Venn diagram shows three events, A, B and C, and their associated probabilities.



Events B and C are mutually exclusive.
Events A and C are independent.

Showing your working, find the value of x, the value of y and the value of z.

(Total for Question 4 is 5 marks)

Gold Questions

Calculator

The total mark for this section is 31

Q1

On a randomly chosen day the probability that Bill travels to school by car, by bicycle or on foot is $\frac{1}{2}$, $\frac{1}{6}$ and $\frac{1}{3}$ respectively. The probability of being late when using these methods of

travel is $\frac{1}{5}$, $\frac{2}{5}$ and $\frac{1}{10}$ respectively.

(a) Draw a tree diagram to represent this information.

(3)

(b) Find the probability that on a randomly chosen day

(i) Bill travels by foot and is late,

(ii) Bill is not late.

(4)

(Total for Question 1 is 7 marks)

Q2

There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,

70 take developing software,

81 take networking,

35 take developing software and systems support,

28 take networking and developing software,

40 take systems support and networking,

4 take all three extra options.

(a) Draw a Venn diagram to represent this information.

(5)

A student from the course is chosen at random.
Find the probability that this student takes

(b) None of the three extra options,

(1)

(c) Networking only.

(1)

(Total for Question 2 is 7 marks)

Q3

A manufacturer carried out a survey of the defects in their soft toys. It is found that the probability of a toy having poor stitching is 0.03 and that a toy with poor stitching has a probability of 0.7 of splitting open. A toy without poor stitching has a probability of 0.02 of splitting open.

(a) Draw a tree diagram to represent this information.

(3)

(b) Find the probability that a randomly chosen soft toy has exactly one of the two defects, poor stitching or splitting open.

(3)

The manufacturer also finds that soft toys can become faded with probability 0.05 and that this defect is independent of poor stitching or splitting open. A soft toy is chosen at random.

(c) Find the probability that the soft toy has none of these 3 defects.

(2)

(d) Find the probability that the soft toy has exactly one of these 3 defects.

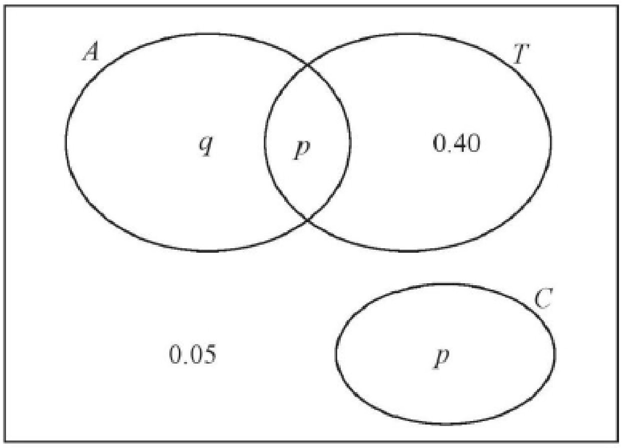
(4)

(Total for Question 3 is 12 marks)

Q4

The Venn diagram shows the probabilities for students at a college taking part in various sports.

> *A* represents the event that a student takes part in Athletics.
> *T* represents the event that a student takes part in Tennis.
> *C* represents the event that a student takes part in Cricket.
> *p* and *q* are probabilities.



The probability that a student selected at random takes part in Athletics or Tennis is 0.75

(a) Find the value of *p*.

(1)

(b) State, giving a reason, whether or not the events *A* and *T* are statistically independent. Show your working clearly.

(3)

(c) Find the probability that a student selected at random does not take part in Athletics or Cricket.

(1)

(Total for Question 4 is 5 marks)

Bronze Questions

Calculator

The total mark for this section is 31

Q1

The probability of a bolt being faulty is 0.3. Find the probability that in a random sample of 20 bolts there are

(a) Exactly 2 faulty bolts,

(2)

(b) More than 3 faulty bolts.

(2)

These bolts are sold in bags of 20. John buys 10 bags.

(c) Find the probability that exactly 6 of these bags contain more than 3 faulty bolts.

(3)

(Total for Question 1 is 7 marks)

Q2

A fair 5-sided spinner has sides numbered 1, 2, 3, 4 and 5

The spinner is spun once and the score of the side it lands on is recorded.

(a) Write down the name of the distribution that can be used to model the score of the side it lands on.

(1)

The spinner is spun 28 times.

The random variable X represents the number of times the spinner lands on 2

(b) (i) Find the probability that the spinner lands on 2 at least 7 times.

(ii) Find $P(4 \leq X < 8)$

(5)

(Total for Question 2 is 6 marks)

Q3

In a game, a player can score 0, 1, 2, 3 or 4 points each time the game is played.

The random variable S, representing the player's score, has the following probability distribution where a, b and c
are constants.

s	0	1	2	3	4
$P(S = s)$	a	b	c	0.1	0.15

The probability of scoring less than 2 points is twice the probability of scoring at least 2 points.

Each game played is independent of previous games played.

John plays the game twice and adds the two scores together to get a total.

Calculate the probability that the total is 6 points.

(Total for Question 3 is 6 marks)

Q4

Naasir is playing a game with two friends. The game is designed to be a game of chance so that the probability of Naasir winning each game is $\frac{1}{3}$

Naasir and his friends play the game 15 times.

(a) Find the probability that Naasir wins

(i) exactly 2 games,
(ii) more than 5 games.

(3)

Naasir claims he has a method to help him win more than $\frac{1}{3}$ of the games. To test this claim, the three of them played the game again 32 times and Naasir won 16 of these games.

(b) Stating your hypotheses clearly, test Naasir's claim at the 5% level of significance.

(4)

(Total for Question 4 is 7 marks)

Q5

A test statistic has a distribution B(25, p).

Given that

$$H_0 : p = 0.5 \quad H_1 : p \neq 0.5$$

(a) Find the critical region for the test statistic such that the probability in each tail is as close as possible to 2.5%.

(3)

(b) State the probability of incorrectly rejecting H_0 using this critical region.

(2)

(Total for Question 5 is 5 marks)

Pearson

Silver Questions

Calculator

The total mark for this section is 34

Q1

Linda regularly takes a taxi to work five times a week. Over a long period of time she finds the taxi is late once a week. The taxi firm changes her driver and Linda thinks the taxi is late more often. In the first week, with the new driver, the taxi is late 3 times.

You may assume that the number of times a taxi is late in a week has a Binomial distribution.

Test, at the 5% level of significance, whether or not there is evidence of an increase in the proportion of times the taxi is late. State your hypotheses clearly.

(Total for Question 1 is 7 marks)

Q2

A potter believes that 20% of pots break whilst being fired in a kiln. Pots are fired in batches of 25.

(a) Let X denote the number of broken pots in a batch. A batch is selected at random. Using a 10% significance level, find the critical region for a two tailed test of the potter's belief. You should state the probability in each tail of your critical region.

(4)

The potter aims to reduce the proportion of pots which break in the kiln by increasing the size of the batch fired. He now fires pots in batches of 50. He then chooses a batch at random and discovers there are 6 pots which broke whilst being fired in the kiln.

(b) Test, at the 5% level of significance, whether or not there is evidence that increasing the number of pots in a batch has reduced the percentage of pots that break whilst being fired in the kiln. State your hypotheses clearly.

(5)

(Total for Question 2 is 9 marks)

Q3

In a manufacturing process 25% of articles are thought to be defective. Articles are produced in batches of 20

(a) A batch is selected at random. Using a 5% significance level, find the critical region for a two tailed test that the probability of an article chosen at random being defective is 0.25
You should state the probability in each tail which should be as close as possible to 0.025

(5)

The manufacturer changes the production process to try to reduce the number of defective articles. She then chooses a batch at random and discovers there are 3 defective articles.

(b) Test at the 5% level of significance whether or not there is evidence that the changes to the process have reduced the percentage of defective articles. State your hypotheses clearly.

(5)

(Total for Question 3 is 10 marks)

Q4

Afrika works in a call centre.

She assumes that calls are independent and knows, from past experience, that on each sales call

that she makes there is a probability of $\frac{1}{6}$ that it is successful.

Afrika makes 9 sales calls.

(a) Calculate the probability that at least 3 of these sales calls will be successful.

(2)

The probability of Afrika making a successful sales call is the same each day.

Afrika makes 9 sales calls on each of 5 different days.

(b) Calculate the probability that at least 3 of the sales calls will be successful on exactly 1 of these days.

(2)

Rowan works in the same call centre as Afrika and believes he is a more successful salesperson.

To check Rowan's belief, Afrika monitors the next 35 sales calls Rowan makes and finds that 11 of the sales calls are successful.

(c) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence to support Rowan's belief.

(4)

(Total for Question 4 is 8 marks)

Gold Questions

Calculator

The total mark for this section is 31

Q1

(a) The discrete random variable $X \sim B(40, 0.27)$

Find $P(X \geq 16)$

(2)

Past records suggest that 30% of customers who buy baked beans from a large supermarket buy them in single tins. A new manager suspects that there has been a change in the proportion of customers who buy baked beans in single tins. A random sample of 20 customers who had bought baked beans was taken.

(b) Write down the hypotheses that should be used to test the manager's suspicion.

(1)

(c) Using a 10% level of significance, find the critical region for a two-tailed test to answer the manager's suspicion. You should state the probability of rejection in each tail, which should be less than 0.05

(3)

(d) Find the actual significance level of a test based on your critical region from part (c).

(1)

One afternoon the manager observes that 12 of the 20 customers who bought baked beans, bought their beans in single tins.

(e) Comment on the manager's suspicion in the light of this observation.

(1)

Later it was discovered that the local scout group visited the supermarket that afternoon to buy food for their camping trip.

(f) Comment on the validity of the model used to obtain the answer to part (e), giving a reason for your answer.

(1)

(Total for Question 1 is 9 marks)

Q2

The proportion of houses in Radville which are unable to receive digital radio is 25%. In a survey of a random sample of 30 houses taken from Radville, the number, X, of houses which are unable to receive digital radio is recorded.

(a) Find $P(5 \leqslant X < 11)$

(3)

A radio company claims that a new transmitter set up in Radville will reduce the proportion of houses which are unable to receive digital radio. After the new transmitter has been set up, a random sample of 15 houses is taken, of which 1 house is unable to receive digital radio.

(b) Test, at the 10% level of significance, the radio company's claim. State your hypotheses clearly.

(5)

(Total for Question 2 is 8 marks)

Q3

Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief.

Julie's hypothesis test is shown below.

$H_0 : p = 0.15$
$H_1 : p \geq 0.15$
Let X = the number of customers who buy chocolate.
$X \sim B(30, 0.15)$
$P(X = 8) = 0.0420$
$0.0420 < 0.05$ so reject H_0
There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test.

(2)

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test.

Give a reason for your answer.

(1)

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05

(2)

(d) Find the actual level of significance of this test.

(1)

(Total for Question 3 is 6 marks)

Q4

A biased spinner can only land on one of the numbers 1, 2, 3 or 4. The random variable X represents the number that the spinner lands on after a single spin and

$P(X = r) = P(X = r + 2)$ for $r = 1, 2$

Given that $P(X = 2) = 0.35$

(a) Find the complete probability distribution of X.

(2)

Ambroh spins the spinner 60 times.

(b) Find the probability that more than half of the spins land on the number 4

Give your answer to 3 significant figures.

(3)

The random variable $Y = \frac{12}{X}$

(c) Find $P(Y - X \leq 4)$

(3)

(Total for Question 4 is 8 marks)