

Question 1

Bob, a gardener, measures the time taken, y minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, x °C. His results are shown in the table.

x	16	20	24	28	32	36	40	44	48	52	56
y	4.7	4.3	3.8	3.5	3.0	2.7	2.4	2.0	1.8	1.6	1.1

- (a) State why the explanatory variable is temperature. (1 mark)
- (b) Calculate the equation of the least squares regression line $y = a + bx$. (4 marks)
- (c) (i) Interpret, in the context of this question, your value for b . (2 marks)
- (ii) Explain why no sensible practical interpretation can be given for your value of a . (2 marks)
- (d) (i) Estimate the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 30 °C. (2 marks)
- (ii) Show why the equation cannot be used to make a valid estimate of the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 75 °C. (2 marks)

Question 2

$P(A) = 0.4$, $P(B) = 0.55$ and $P(C) = 0.26$.

Given that $P(A \cap B) = 0.2$, that events A and C are mutually exclusive and that events B and C are statistically independent,

- a** Draw a Venn diagram to illustrate events A , B and C . (5)
- b** Show that events A and B are not statistically independent. (2)
- c** Find $P(A|B')$. (2)
- d** Find $P(C|(A \cap B)')$. (2)

Question 3

An industrial chemical process produces an amount of a substance, q grams, dependent on the temperature, t °C applied. The table below shows the outcomes of five experiments.

t	10	20	32	41	57
q	160	700	2000	3300	6400

A chemist believes that the relationship between the variables can be modelled by an equation of the form $q = kt^n$, where k and n are constants to be determined. The data are coded using $x = \log t$ and $y = \log q$. The product moment correlation coefficient between x and y is found to be 0.9998.

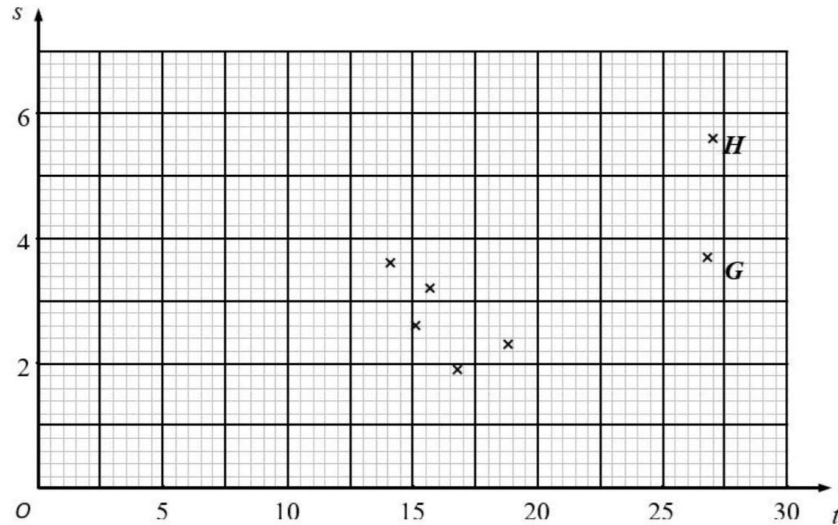
- a** State with a reason whether this value supports the suggested model. **(1)**
- b** Given that the equation of the regression line of y on x is $y = 0.0761 + 2.1317x$, find the value of k and the value of n . **(3)**
- c** Explain, giving a reason, whether it would be sensible to use this model to predict the amount of substance produced when $t = 85$ °C. **(1)**

Question 4

Jemima plays two games of tennis. The probability that she wins the first game is 0.62. If she wins the first game, the probability that she wins the second is 0.75. If she loses the first game, the probability that she wins the second is 0.45. Find the probability that she wins both games given that she wins the second game. **(4)**

Question 5

2. A researcher believes that there is a linear relationship between daily mean temperature and daily total rainfall. The 7 places in the northern hemisphere from the large data set are used. The mean of the daily mean temperatures, $t^{\circ}\text{C}$, and the mean of the daily total rainfall, s mm, for the month of July in 2015 are shown on the scatter diagram below.



- (a) With reference to the scatter diagram, explain why a linear regression model may not be suitable for the relationship between t and s . (1)

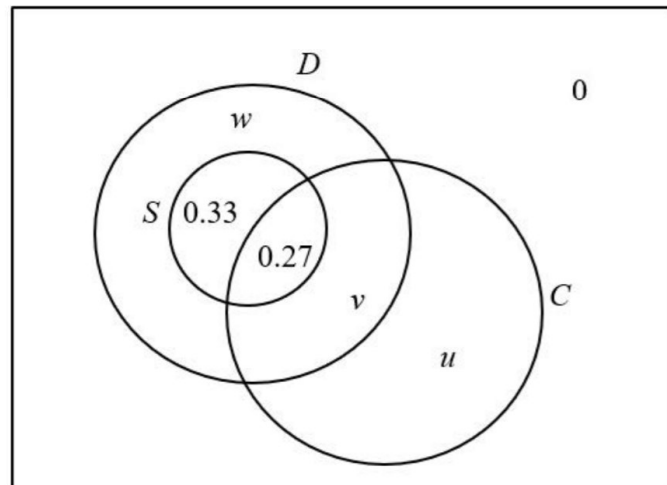
The researcher calculated the product moment correlation coefficient for the 7 places and obtained $r = 0.658$

- (b) Stating your hypotheses clearly, test at the 10% level of significance, whether or not the product moment correlation coefficient for the population is greater than zero. (3)
- (c) Using your knowledge of the large data set, suggest the names of the 2 places labelled G and H . (1)
- (d) Using your knowledge from the large data set, and with reference to the locations of the 2 places labelled G and H , give a reason why these places have the highest temperatures in July. (1)
- (e) Suggest how you could make better use of the large data set to investigate the relationship between daily mean temperature and daily total rainfall. (1)

Question 6

The Venn diagram shows the probabilities of students' lunch boxes containing a drink, sandwiches and a chocolate bar.

D is the event that a lunch box contains a drink,
 S is the event that a lunch box contains sandwiches,
 C is the event that a lunch box contains a chocolate bar,
 u , v and w are probabilities.



- (a) Write down $P(S \cap D')$ (1)

One day, 80 students each bring in a lunch box.
Given that all 80 lunch boxes contain sandwiches and a drink,

- (b) estimate how many of these 80 lunch boxes will contain a chocolate bar. (3)

Given that the events S and C are independent and that $P(D|C) = \frac{14}{15}$

- (c) calculate the value of u , the value of v and the value of w . (7)

Question 7

An ornithologist believes that there is a relationship between the tail length, t mm, and the wing length, w mm, of female hook-billed kites. A random sample of size 10 is taken from a database of these kites and the relevant data is given in the table below.

t (mm)	191	197	208	180	188	210	196	191	179	208
w (mm)	284	285	288	273	280	283	288	271	257	289

The ornithologist plans to use a linear regression model based on these data and interpolate or extrapolate as necessary to estimate the wing length of other female hook-billed kites from their tail length.

(a) (i) Explain what is meant by extrapolation. (1)

(ii) Explain the dangers of extrapolation. (1)

The ornithologist attempts to calculate the product moment correlation coefficient, r , and obtains a value of 1.3

(b) Explain how she should be able to identify that this is incorrect without carrying out any further calculations. (1)

(c) Use your calculator to find the correct value of the product moment correlation coefficient, r . (1)

(d) Stating your hypotheses clearly test, at the 1% significance level, whether or not there is evidence that the product moment correlation coefficient for the population is positive. (3)

(e) Explain what your test in part (d) suggests about female hook-billed kites. (1)

Question 8

A company maintains machines.

It has two types of contract, a service contract and a repair contract.

The company classes its customers as new customers or existing customers.

The table gives information about the company's customers.

	Service contract	Repair contract
New customer	65	82
Existing customer	231	262

The company is going to survey its customers. It plans to take a sample of 100 of its customers, stratified by customer type and contract type.

(a) Work out how many new customers with repair contracts should be sampled.

(2)

The company has developed a test for a certain fault in the machines it services.

The test sometimes gives incorrect results.

The company collects information from a sample of randomly selected machines.

- 2% of the machines have the fault
- 70% of the machines with the fault test positive for the fault
- 10% of the machines without the fault test positive for the fault.

A machine is selected at random from the sample of the machines, and tests positive for the fault.

(b) (i) Calculate the probability that the machine has the fault.

(4)

(ii) Comment on the usefulness of the company's test.

Give a reason for your answer.

(1)

When the company services the machines, it checks two components, α and β , for wear and tear and replaces these if needed.

Event A is that component α needs to be replaced.

Event B is that component β needs to be replaced.

The probability that component α needs to be replaced is 0.35

The probability that component β needs to be replaced is 0.55

The probability that neither component needs to be replaced is 0.28

(c) Show that events A and B are not independent.

(2)

(d) Find the probability that component α or component β needs to be replaced, but not both.

(2)

Question 9

Answer ALL questions

- 1 As part of a study of charity shops in a small market town, two such shops, X and Y , were each asked to provide details of its takings on 12 randomly selected days.

The table shows, for each of the 12 days, the day's takings, £ x , of charity shop X and the day's takings, £ y , of charity shop Y .

Day	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>
x	46	57	39	116	62	77	41	61	15	53	68	61
y	78	102	66	214	98	72	98	134	21	67	95	83

Table 1

- (a) Find the value of the product moment correlation coefficient between x and y and perform a test to investigate whether there is significant correlation between the day's takings at Shops X and Y .

You must state any assumptions you have made.

(7)

Alex, the person carrying out this study realised, subsequently, that one of the 12 selected days was a particularly popular town market day and another was a day on which the weather was extremely severe.

- (b) Identify **each** of these two days, giving a **reason** for each choice.

(2)

Alex decides to remove the data for these two days and then he correctly recalculates the product moment correlation coefficient as $r = 0.1827$.

- (c) Revise the hypothesis test carried out in part (a) using Alex's recalculated product moment correlation coefficient.

(2)

- (d) Comment on Alex's decision to remove the data for these two days, and discuss the validity of the conclusion reached in part (c).

(4)

(Total for Question 1 is 15 marks)

Question 10

New-born babies are tested for a particular medical condition. If this condition is detected promptly it can be treated, but otherwise permanent damage occurs.

About 1 baby in 10 000 is born with this condition. A test is needed which detects the condition as reliably as possible.

For one of the available tests, the probability of detecting the condition in a new-born baby who is affected is 99.5%.

However, there is also a probability of 0.4% that the test indicates the condition when in fact the new-born baby is not affected.

- (a) What is the probability that a test on a randomly chosen new-born baby will indicate the presence of the condition? (2)

Sarah and Abraham have just heard that the test on their new-born baby indicates the presence of the condition.

- (b) What is the probability that their baby really does have the condition? (3)

Using the information about this test, a newspaper runs the following headline:

“NEW TEST FOR INFANT CONDITION FOUND TO BE 99.5% ACCURATE!”

- (c) Give **two** reasons to doubt the validity of this statement. (2)

(Total for Question 2 is 7 marks)