Topic 7 Part 2 *[433 marks]*

The random variable X has a geometric distribution with parameter p.

1a. Show that
$P(X \leqslant n) = 1 - (1-p)^n$, $n \in \mathbb{Z}^+$.

[3 marks]

1b. Deduce an expression for
$P(m < X \leqslant n)$, m, $n \in \mathbb{Z}^+$ and $m < n$.

[1 mark]

1c. Given that $p = 0.2$, find the least value of n for which
$P(1 < X \leqslant n) > 0.5$, $n \in \mathbb{Z}^+$.

[2 marks]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} 2x, & 0 \leqslant x \leqslant 0.5, \\ \frac{4}{3} - \frac{2}{3}x, & 0.5 \leqslant x \leqslant 2 \\ 0, & \text{otherwise.} \end{cases}$$

2a. Sketch the function f and show that the lower quartile is 0.5.

[3 marks]

2b. (i)   Determine E(X).
(ii)   Determine
E(X^2).

[4 marks]

2c. Two independent observations are made from X and the values are added.
The resulting random variable is denoted Y.
(i)   Determine
E($Y - 2X$) .
(ii)   Determine
Var$(Y - 2X)$.

[5 marks]

2d. (i)   Find the cumulative distribution function for X .
(ii)   Hence, or otherwise, find the median of the distribution.

[7 marks]

The weights of the oranges produced by a farm may be assumed to be normally distributed with mean 205 grams and standard deviation 10 grams.

3a. Find the probability that a randomly chosen orange weighs more than 200 grams.

[2 marks]

3b. Five of these oranges are selected at random to be put into a bag. Find the probability that the combined weight of the five oranges is less than 1 kilogram.

[4 marks]

3c. The farm also produces lemons whose weights may be assumed to be normally distributed with mean 75 grams and standard deviation 3 grams. Find the probability that the weight of a randomly chosen orange is more than three times the weight of a randomly chosen lemon.

[5 marks]

Ten friends try a diet which is claimed to reduce weight. They each weigh themselves before starting the diet, and after a month on the diet, with the following results.

Friend	A	B	C	D	E	F	G	H	I	J
Weight before (kg)	68.4	70.9	74.7	65.4	59.4	69.0	73.9	62.6	68.3	58.2
Weight after (kg)	66.2	67.4	70.4	65.9	55.2	69.2	71.4	59.9	68.2	58.9

4a. Determine unbiased estimates of the mean and variance of the loss in weight achieved over the month by people using this diet. *[5 marks]*

4b.
(i)   State suitable hypotheses for testing whether or not this diet causes a mean loss in weight. *[6 marks]*

(ii)   Determine the value of a suitable statistic for testing your hypotheses.

(iii)   Find the 1 % critical value for your statistic and state your conclusion.

The random variable X has a Poisson distribution with unknown mean

μ . It is required to test the hypotheses

$H_0 : \mu = 3$ against

$H_1 : \mu \neq 3$.

Let S denote the sum of 10 randomly chosen values of X . The critical region is defined as

$(S \leqslant 22) \cup (S \geqslant 38)$.

5a. Calculate the significance level of the test. *[5 marks]*

5b. Given that the value of *[5 marks]*

μ is actually 2.5, determine the probability of a Type II error.

6a. The random variable X has the negative binomial distribution $\mathrm{NB}(3, p)$. *[7 marks]*

Let

$f(x)$ denote the probability that X takes the value x .

(i)   Write down an expression for

$f(x)$, and show that

$$\ln f(x) = 3\ln\left(\frac{p}{1-p}\right) + \ln(x-1) + \ln(x-2) + x\ln(1-p) - \ln 2 .$$

(ii)   State the domain of f .

(iii)   The domain of f is extended to

$]2, \infty[$. Show that

$\frac{f'(x)}{f(x)} = \frac{1}{x-1} + \frac{1}{x-2} + \ln(1-p)$.

6b. Jo has a biased coin which has a probability of 0.35 of showing heads when tossed. She tosses this coin successively and the 3rd head occurs on the Yth toss. Use the result in part (a)(iii) to find the most likely value of Y.

7. The mean weight of a certain breed of bird is believed to be 2.5 kg. In order to test this belief, it is planned to determine the weights x_1, x_2, x_3, ..., x_{16} (in kg) of sixteen of these birds and then to calculate the sample mean $\bar{x}$. You may assume that these weights are a random sample from a normal distribution with standard deviation 0.1 kg.

(a) State suitable hypotheses for a two-tailed test.

(b) Find the critical region for $\bar{x}$ having a significance level of 5 %.

(c) Given that the mean weight of birds of this breed is actually 2.6 kg, find the probability of making a Type II error.

8a. Alan and Brian are athletes specializing in the long jump. When Alan jumps, the length of his jump is a normally distributed random variable with mean 5.2 metres and standard deviation 0.1 metres. When Brian jumps, the length of his jump is a normally distributed random variable with mean 5.1 metres and standard deviation 0.12 metres. For both athletes, the length of a jump is independent of the lengths of all other jumps. During a training session, Alan makes four jumps and Brian makes three jumps. Calculate the probability that the mean length of Alan's four jumps is less than the mean length of Brian's three jumps.

8b. Colin joins the squad and the coach wants to know the mean length, μ metres, of his jumps. Colin makes six jumps resulting in the following lengths in metres.

5.21, 5.30, 5.22, 5.19, 5.28, 5.18

(i) Calculate an unbiased estimate of both the mean μ and the variance of the lengths of his jumps.

(ii) Assuming that the lengths of these jumps are independent and normally distributed, calculate a 90 % confidence interval for μ.

9. Anna cycles to her new school. She records the times taken for the first ten days with the following results (in minutes).

12.4 13.7 12.5 13.4 13.8 12.3 14.0 12.8 12.6 13.5

Assume that these times are a random sample from the $N(\mu, \sigma^2)$ distribution.

(a) Determine unbiased estimates for μ and σ^2.

(b) Calculate a 95 % confidence interval for μ.

(c) Before Anna calculated the confidence interval she thought that the value of μ would be 12.5. In order to check this, she sets up the null hypothesis $H_0 : \mu = 12.5$.

(i) Use the above data to calculate the value of an appropriate test statistic. Find the corresponding p-value using a two-tailed test.

(ii) Interpret your p-value at the 1 % level of significance, justifying your conclusion.

10. The random variable X has a Poisson distribution with mean

    μ. The value of

    μ is known to be either 1 or 2 so the following hypotheses are set up.

    $$H_0 : \mu = 1; \ H_1 : \mu = 2$$

    A random sample

    $x_1, x_2, \ldots, x_{10}$ of 10 observations is taken from the distribution of X and the following critical region is defined.

    $$\sum_{i=1}^{10} x_i \geqslant 15$$

    Determine the probability of

    (a)   a Type I error;

    (b)   a Type II error.

11. A shop sells apples, pears and peaches. The weights, in grams, of these three types of fruit may be assumed to be normally distributed with means and standard deviations as given in the following table.

Fruit	Mean	Standard Deviation
Apples	115	5
Pears	110	4
Peaches	105	3

    Alan buys 1 apple and 1 pear while Brian buys 1 peach. Calculate the probability that the combined weight of Alan's apple and pear is greater than twice the weight of Brian's peach.

12. The random variable X has the negative binomial distribution NB$(5, p)$, where $p < 0.5$, and

    $P(X = 10) = 0.05$. By first finding the value of p, find the value of

    $P(X = 11)$.

13. The company *Fresh Water* produces one-litre bottles of mineral water. The company wants to determine the amount of

    magnesium, in milligrams, in these bottles.

    A random sample of ten bottles is analysed and the results are as follows:

    $$6.7, 7.2, 6.7, 6.8, 6.9, 7.0, 6.8, 6.6, 7.1, 7.3.$$

    Find unbiased estimates of the mean and variance of the amount of magnesium in the one-litre bottles.

14. A hospital specializes in treating overweight patients. These patients have weights that are independently, normally distributed

    with mean 200 kg and standard deviation 15 kg. The elevator in the hospital will break if the total weight of people inside it exceeds 1150 kg. Six patients enter the elevator.

    Find the probability that the elevator breaks.

15. The length of time, T, in months, that a football manager stays in his job before he is removed can be approximately modelled
by a normal distribution with population mean

μ and population variance

σ^2. An independent sample of five values of T is given below.

6.5, 12.4, 18.2, 3.7, 5.4

(a)   Given that

$\sigma^2 = 9$,

(i)   use the above sample to find the 95 % confidence interval for

μ, giving the bounds of the interval to two decimal places;

(ii)   find the smallest number of values of T that would be required in a sample for the total width of the 90 % confidence interval for

μ to be less than 2 months.

(b)   If the value of

σ^2 is unknown, use the above sample to find the 95 % confidence interval for

μ, giving the bounds of the interval to two decimal places.

16. As soon as Sarah misses a total of 4 lessons at her school an email is sent to her parents. The probability that she misses any
particular lesson is constant with a value of

$\frac{1}{3}$. Her decision to attend a lesson is independent of her previous decisions.

(a)   Find the probability that an email is sent to Sarah's parents after the

8th lesson that Sarah was scheduled to attend.

(b)   If an email is sent to Sarah's parents after the

X^{th} lesson that she was scheduled to attend, find

$E(X)$.

(c)   If after 6 of Sarah's scheduled lessons we are told that she has missed exactly 2 lessons, find the probability that an email is sent to

her parents after a total of 12 scheduled lessons.

(d)   If we know that an email was sent to Sarah's parents immediately after her

6th scheduled lesson, find the probability that Sarah missed her

2nd scheduled lesson.

17. A teacher has forgotten his computer password. He knows that it is either six of the letter J followed by two of the letter R (*i.e.*
JJJJJJRR) or three of the letter J followed by four of the letter R (*i.e.* JJJRRRR). The computer is able to tell him at random just two of the

letters in his password.

The teacher decides to use the following rule to attempt to find his password.

If the computer gives him a J and a J, he will accept the null hypothesis that his password is JJJJJJRR.

Otherwise he will accept the alternative hypothesis that his password is JJJRRRR.

(a)   Define a Type I error.

(b)   Find the probability that the teacher makes a Type I error.

(c)   Define a Type II error.

(d)   Find the probability that the teacher makes a Type II error.

The weight of tea in *Supermug* tea bags has a normal distribution with mean 4.2 g and standard deviation 0.15 g. The weight of tea in *Megamug* tea bags has a normal distribution with mean 5.6 g and standard deviation 0.17 g.

18a. Find the probability that a randomly chosen *Supermug* tea bag contains more than 3.9 g of tea.

[2 marks]

18b. Find the probability that, of two randomly chosen *Megamug* tea bags, one contains more than 5.4 g of tea and one contains

[4 marks]

less than 5.4 g of tea.

18c. Find the probability that five randomly chosen *Supermug* tea bags contain a total of less than 20.5 g of tea.

[4 marks]

18d. Find the probability that the total weight of tea in seven randomly chosen *Supermug* tea bags is more than the total weight in

[5 marks]

five randomly chosen *Megamug* tea bags.

The random variable X represents the lifetime in hours of a battery. The lifetime may be assumed to be a continuous random variable

X with a probability density function given by

$f(x) = \lambda e^{-\lambda x}$, where

$x \geqslant 0$.

19a. Find the cumulative distribution function,

[3 marks]

$F(x)$, of X.

19b. Find the probability that the lifetime of a particular battery is more than twice the mean.

[2 marks]

19c. Find the median of X in terms of

[3 marks]

λ.

19d. Find the probability that the lifetime of a particular battery lies between the median and the mean.

[2 marks]

The cumulative frequency graph below represents the weight in grams of 80 apples picked from a particular tree.
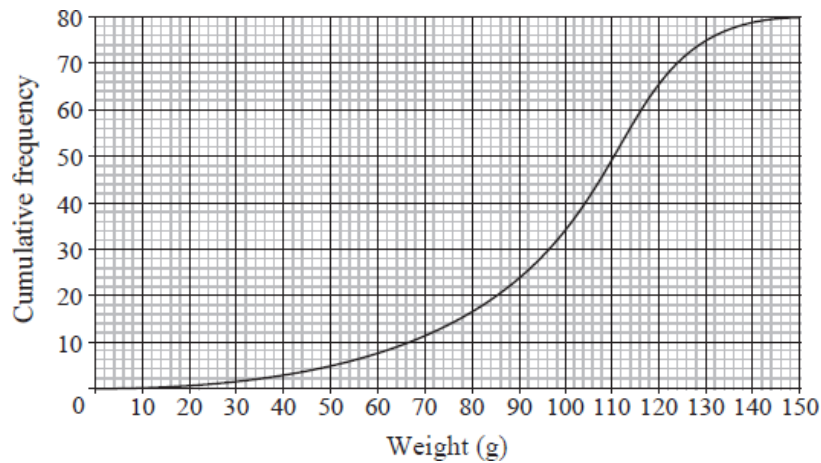
20a. Estimate the [2 marks]

(i)   median weight of the apples;

(ii) 30th percentile of the weight of the apples.

20b. Estimate the number of apples which weigh more than 110 grams. [2 marks]

21. A coin was tossed 200 times and 115 of these tosses resulted in 'heads'. Use a two-tailed test with significance level 1 % to [8 marks] investigate whether or not the coin is biased.

A shop sells apples and pears. The weights, in grams, of the apples may be assumed to have a $N(200, 15^2)$ distribution and the weights of the pears, in grams, may be assumed to have a $N(120, 10^2)$ distribution.

22. (a)   Find the probability that the weight of a randomly chosen apple is more than double the weight of a randomly chosen [14 marks] pear.

(b)   A shopper buys 3 apples and 4 pears. Find the probability that the total weight is greater than 1000 grams.

The random variable X is normally distributed with unknown mean

μ and unknown variance

σ^2. A random sample of 10 observations on X was taken and the following 95 % confidence interval for

μ was correctly calculated as [4.35, 4.53] .

23. (a)   Calculate an unbiased estimate for [14 marks]

(i)

μ,

(ii)

σ^2.

(b)   The value of

μ is thought to be 4.5, so the following hypotheses are defined.

$$H_0 : \mu = 4.5; \ H_1 : \mu < 4.5$$

(i)   Find the p-value of the observed sample mean.

(ii) State your conclusion if the significance level is

(a)   1 %,

(b)   10 %.

Anna has a fair cubical die with the numbers 1, 2, 3, 4, 5, 6 respectively on the six faces. When she tosses it, the score is defined as the number on the uppermost face. One day, she decides to toss the die repeatedly until all the possible scores have occurred at least once.

24. (a) Having thrown the die once, she lets [10 marks]

X_2 denote the number of additional throws required to obtain a different number from the one obtained on the first throw. State the distribution of

X_2 and hence find

$E(X_2)$.

(b) She then lets

X_3 denote the number of additional throws required to obtain a different number from the two numbers already obtained. State the distribution of

X_3 and hence find

$E(X_3)$.

(c) By continuing the process, show that the expected number of tosses needed to obtain all six possible scores is 14.7.

25a. The random variable Y is such that [6 marks]

$E(2Y+3) = 6$ and $\mathrm{Var}(2-3Y) = 11$.

Calculate

(i) $E(Y)$;

(ii)

$\mathrm{Var}(Y)$;

(iii)

$E(Y^2)$.

25b. Independent random variables R and S are such that [6 marks]

$$R \sim N(5,\ 1) \text{ and } S \sim N(8,\ 2).$$

The random variable V is defined by $V = 3S - 4R$.

Calculate $P(V > 5)$.

26. A factory makes wine glasses. The manager claims that on average 2 % of the glasses are imperfect. A random sample of 200 [7 marks] glasses is taken and 8 of these are found to be imperfect.

Test the manager's claim at a 1 % level of significance using a one-tailed test.

A teacher wants to determine whether practice sessions improve the ability to memorize digits.

He tests a group of 12 children to discover how many digits of a twelve-digit number could be repeated from memory after hearing them once. He gives them test 1, and following a series of practice sessions, he gives them test 2 one week later. The results are shown in the table below.

Child	A	B	C	D	E	F	G	H	I	J	K	L
Number of digits remembered on test 1	4	6	4	7	8	5	6	7	6	8	4	7
Number of digits remembered on test 2	7	8	5	5	10	7	7	10	8	6	3	9

[11 marks]

27. (a) State appropriate null and alternative hypotheses.

   (b) Test at the 5 % significance level whether or not practice sessions improve ability to memorize digits, justifying your choice of test.

A population is known to have a normal distribution with a variance of 3 and an unknown mean μ. It is proposed to test the hypotheses

$H_0 : \mu = 13$, $H_1 : \mu > 13$ using the mean of a sample of size 2.

[16 marks]

28. (a) Find the appropriate critical regions corresponding to a significance level of

   (i)   0.05;

   (ii) 0.01.

   (b) Given that the true population mean is 15.2, calculate the probability of making a Type II error when the level of significance is

   (i)   0.05;

   (ii) 0.01.

   (c) How is the change in the probability of a Type I error related to the change in the probability of a Type II error?

[9 marks]

29. The apple trees in a large orchard have, for several years, suffered from a disease for which the outward sign is a red discolouration on some leaves.

The fruit grower knows that the mean number of discoloured leaves per tree is 42.3. The fruit grower suspects that the disease is caused by an infection from a nearby group of cedar trees. He cuts down the cedar trees and, the following year, counts the number of discoloured leaves on a random sample of seven apple trees. The results are given in the table below.

Tree	A	B	C	D	E	F	G
Number of red leaves	32	16	57	28	55	12	45

   (a) From these data calculate an unbiased estimate of the population variance.

   (b) Stating null and alternative hypotheses, carry out an appropriate test at the 10 % level to justify the cutting down of the cedar trees.

[12 marks]

30. (a) The heating in a residential school is to be increased on the third frosty day during the term. If the probability that a day will be frosty is 0.09, what is the probability that the heating is increased on the 25$^{\text{th}}$ day of the term?

   (b) On which day is the heating most likely to be increased?

31.

(a) A random variable, X, has probability density function defined by

$$f(x) = \begin{cases} 100, & \text{for } -0.005 \leqslant x < 0.005 \\ 0, & \text{otherwise.} \end{cases}$$

Determine $E(X)$ and $\text{Var}(X)$.

(b) When a real number is rounded to two decimal places, an error is made.

Show that this error can be modelled by the random variable X.

(c) A list contains 20 real numbers, each of which has been given to two decimal places. The numbers are then added together.

(i) Write down bounds for the resulting error in this sum.

(ii) Using the central limit theorem, estimate to two decimal places the probability that the absolute value of the error exceeds 0.01.

(iii) State clearly any assumptions you have made in your calculation.

32. Ahmed and Brian live in the same house. Ahmed always walks to school and Brian always cycles to school. The times taken to travel to school may be assumed to be independent and normally distributed. The mean and the standard deviation for these times are shown in the table below.

	Mean (minutes)	Standard Deviation (minutes)
Ahmed Walking	30	3
Brian Cycling	12	2

(a) Find the probability that on a particular day Ahmed takes more than 35 minutes to walk to school.

(b) Brian cycles to school on five successive mornings. Find the probability that the total time taken is less than 70 minutes.

(c) Find the probability that, on a particular day, the time taken by Ahmed to walk to school is more than twice the time taken by Brian to cycle to school.

33.

(a) After a chemical spillage at sea, a scientist measures the amount, x units, of the chemical in the water at 15 randomly chosen sites. The results are summarised in the form

$\sum x = 18$ and

$\sum x^2 = 28.94$. Before the spillage occurred the mean level of the chemical in the water was 1.1. Test at the 5 % significance level the hypothesis that there has been an increase in the amount of the chemical in the water.

(b) Six months later the scientist returns and finds that the mean amount of the chemical in the water at the 15 randomly chosen sites is 1.18. Assuming that this sample came from a normal population with variance 0.0256, find a 90 % confidence interval for the mean level of the chemical.

34. In a game there are *n* players, where

$n > 2$. Each player has a disc, one side of which is red and one side blue. When thrown, the disc is equally likely to show red or blue. All players throw their discs simultaneously. A player wins if his disc shows a different colour from all the other discs. Players throw repeatedly until one player wins.

Let *X* be the number of throws each player makes, up to and including the one on which the game is won.

(a) State the distribution of *X* .

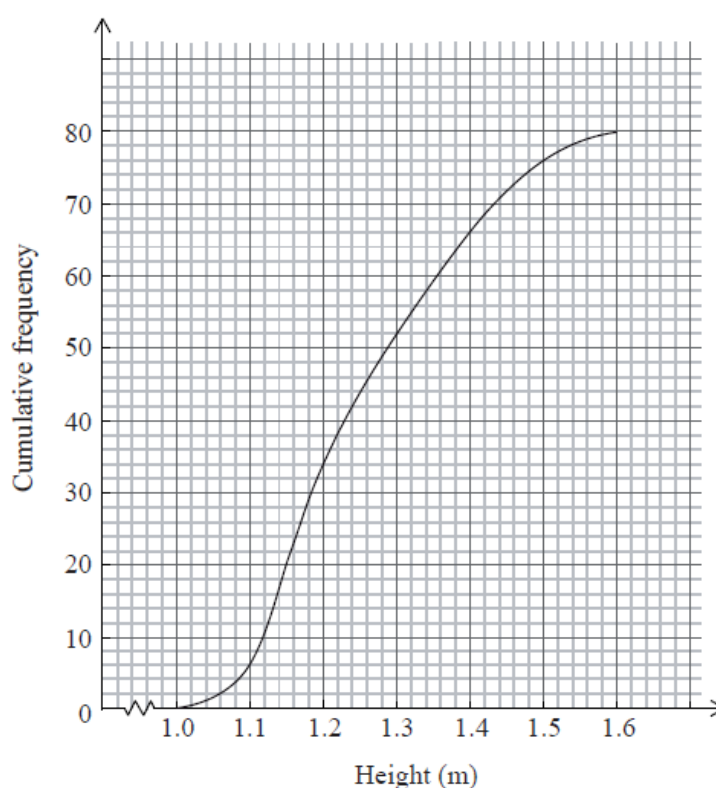(b) Find

$P(X = x)$ in terms of *n* and *x* .

(c) Find

$E(X)$ in terms of *n* .

(d) Given that $n = 7$, find the least number, *k* , such that

$P(X \leqslant k) > 0.5$.

The heights of all the new boys starting at a school were measured and the following cumulative frequency graph was produced.

35a. Complete the grouped frequency table for these data.

Interval	Frequency
]1.0, 1.1]	
]1.1, 1.2]	
]1.2, 1.3]	
]1.3, 1.4]	
]1.4, 1.5]	
]1.5, 1.6]	

35b. Estimate the mean and standard deviation of the heights of these 80 boys.

35c. Explain briefly whether or not the normal distribution provides a suitable model for this population. *[2 marks]*

A shopper buys 12 apples from a market stall and weighs them with the following results (in grams).

$$117, 124, 129, 118, 124, 116, 121, 126, 118, 121, 122, 129$$

You may assume that this is a random sample from a normal distribution with mean

μ and variance

σ^2.

36a. Determine unbiased estimates of *[3 marks]*

μ and

σ^2.

36b. Determine a 99 % confidence interval for *[2 marks]*

μ .

36c. The stallholder claims that the mean weight of apples is 125 grams but the shopper claims that the mean is less than this. *[5 marks]*

(i)   State suitable hypotheses for testing these claims.

(ii)   Calculate the *p*-value of the above sample.

(iii) Giving a reason, state which claim is supported by your *p*-value using a 5 % significance level.

When Andrew throws a dart at a target, the probability that he hits it is

$\frac{1}{3}$; when Bill throws a dart at the target, the probability that he hits the it is

$\frac{1}{4}$. Successive throws are independent. One evening, they throw darts at the target alternately, starting with Andrew, and stopping as soon as one of their darts hits the target. Let X denote the total number of darts thrown.

37a. Write down the value of *[2 marks]*

$P(X = 1)$ and show that
$P(X = 2) = \frac{1}{6}$.

37b. Show that the probability generating function for X is given by *[6 marks]*

$$G(t) = \frac{2t + t^2}{6 - 3t^2}.$$

37c. Hence determine *[4 marks]*

$E(X)$.

The weights of adult monkeys of a certain species are known to be normally distributed, the males with mean 30 kg and standard deviation 3 kg and the females with mean 20 kg and standard deviation 2.5 kg.

38a. Find the probability that the weight of a randomly selected male is more than twice the weight of a randomly selected female. *[5 marks]*

38b. Two males and five females stand together on a weighing machine. Find the probability that their total weight is less than 175 *[4 marks]*

    kg.

The students in a class take an examination in Applied Mathematics which consists of two papers. Paper 1 is in Mechanics and Paper

2 is in Statistics. The marks obtained by the students in Paper 1 and Paper 2 are denoted by

(x, y) respectively and you may assume that the values of

(x, y) form a random sample from a bivariate normal distribution with correlation coefficient

ρ. The teacher wishes to determine whether or not there is a positive association between marks in Mechanics and marks in Statistics.

39a. State suitable hypotheses.   *[1 mark]*

                                                                                                                 [5 marks]
39b. The marks obtained by the 12 students who sat both papers are given in the following table.

Student	A	B	C	D	E	F	G	H	I	J	K	L
x	52	47	82	69	38	50	72	46	23	60	42	53
y	55	44	79	62	41	37	71	44	31	45	47	49

(i)   Determine the product moment correlation coefficient for these data and state its p-value.

(ii)   Interpret your p-value in the context of the problem.

39c. George obtained a mark of 63 on Paper 1 but was unable to sit Paper 2 because of illness. Predict the mark that he would have *[4 marks]*

    obtained on Paper 2.

                                                                                                                 [5 marks]
39d. Another class of 16 students sat examinations in Physics and Chemistry and the product moment correlation coefficient

    between the marks in these two subjects was calculated to be 0.524. Using a 1 % significance level, determine whether or not this value

    suggests a positive association between marks in Physics and marks in Chemistry.