

APOLLO

Statistics and probability

HL

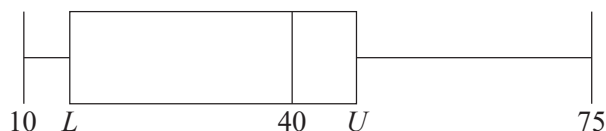
Mathematics: analysis and approaches

First assessment 2021

87 questions | questions

3. [Maximum mark: 5]

diagram not to scale



(a) Find the minimum possible value of U . [3]

(b) Hence, find the minimum possible value of L . [2]

[illegible]

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 16]

A biased four-sided die, A, is rolled. Let X be the score obtained when die A is rolled. The probability distribution for X is given in the following table.

x	1	2	3	4
$P(X=x)$	p	p	p	$\frac{1}{2}p$

(a) Find the value of p . [2]

(b) Hence, find the value of $E(X)$. [2]

A second biased four-sided die, B, is rolled. Let Y be the score obtained when die B is rolled. The probability distribution for Y is given in the following table.

y	1	2	3	4
$P(Y=y)$	q	q	q	r

(c) (i) State the range of possible values of r .

(ii) Hence, find the range of possible values of q . [3]

(d) Hence, find the range of possible values for $E(Y)$. [3]

Agnes and Barbara play a game using these dice. Agnes rolls die A once and Barbara rolls die B once. The probability that Agnes' score is less than Barbara's score is $\frac{1}{2}$.

(e) Find the value of $E(Y)$. [6]



Turn over

Do **not** write solutions on this page.

Section A

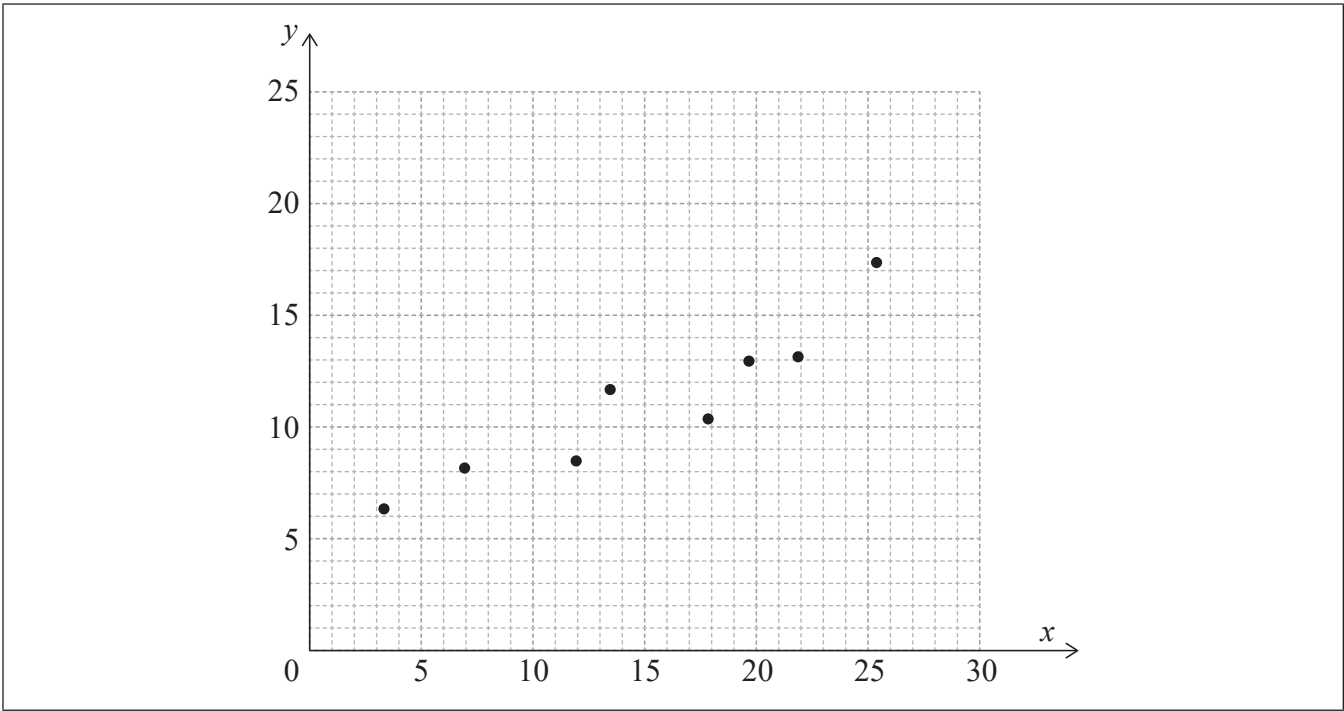
Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

1. [Maximum mark: 7]

The following table shows the data collected from an experiment.

x	3.3	6.9	11.9	13.4	17.8	19.6	21.8	25.3
y	6.3	8.1	8.4	11.6	10.3	12.9	13.1	17.3

The data is also represented on the following scatter diagram.



(This question continues on the following page)



(Question 1 continued)

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$, where $a, b \in \mathbb{R}$.

- (a) Write down the value of a and the value of b . [2]
- (b) Use this model to predict the value of y when $x = 18$. [2]
- (c) Write down the value of $\bar{x}$ and the value of $\bar{y}$. [1]
- (d) Draw the line of best fit on the scatter diagram. [2]



This image shows a full page of dot grid paper. It consists of multiple horizontal rows of small, evenly spaced black dots on a white background. The dots are arranged in straight lines across the width of the page, providing a guide for writing or drawing without solid lines.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

6. [Maximum mark: 6]

A continuous random variable X has the probability density function f_n given by

$$f_n(x) = \begin{cases} (n+1)x^n, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

where $n \in \mathbb{R}, n \geq 0$.

(a) Show that $E(X) = \frac{n+1}{n+2}$. [2]

(b) Show that $\text{Var}(X) = \frac{n+1}{(n+2)^2(n+3)}$. [4]



Turn over

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 15]

The flight times, T minutes, between two cities can be modelled by a normal distribution with a mean of 75 minutes and a standard deviation of σ minutes.

- (a) Given that 2% of the flight times are longer than 82 minutes, find the value of σ . [3]
- (b) Find the probability that a randomly selected flight will have a flight time of more than 80 minutes. [2]
- (c) Given that a flight between the two cities takes longer than 80 minutes, find the probability that it takes less than 82 minutes. [4]

On a particular day, there are 64 flights scheduled between these two cities.

- (d) Find the expected number of flights that will have a flight time of more than 80 minutes. [3]
- (e) Find the probability that more than 6 of the flights on this particular day will have a flight time of more than 80 minutes. [3]



Turn over

Do **not** write solutions on this page.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

1. [Maximum mark: 7]

In Lucy's music academy, eight students took their piano diploma examination and achieved scores out of 150. For her records, Lucy decided to record the average number of hours per week each student reported practising in the weeks prior to their examination. These results are summarized in the table below.

Average weekly practice time (h)	28	13	45	33	17	29	39	36
Diploma score (D)	115	82	120	116	79	101	110	121

- (a) Find Pearson's product-moment correlation coefficient, r , for these data. [2]
- (b) The relationship between the variables can be modelled by the regression equation $D = ah + b$. Write down the value of a and the value of b . [1]
- (c) One of these eight students was disappointed with her result and wished she had practised more. Based on the given data, determine how her score could have been expected to alter had she practised an extra five hours per week. [2]
- (d) Lucy asserts that the number of hours a student practises has a direct effect on their final diploma result. Comment on the validity of Lucy's assertion. [1]

Lucy suspected that each student had not been practising as much as they reported. In order to compensate for this, Lucy deducted a fixed number of hours per week from each of the students' recorded hours.

- (e) State how, if at all, the value of r would be affected. [1]

(This question continues on the following page)



(Question 1 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

— T —

00217102

Please **do not** write on this page.

Answers written on this page
will not be marked.



7. [Maximum mark: 6]

$$f(x) = \begin{cases} \arccos x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine the value of m . [2]

(b) Given that $P(|X - m| \leq a) = 0.3$, determine the value of a . [4]

This image shows a full page of white paper with horizontal dotted lines, typical of primary school handwriting practice paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

4. [Maximum mark: 7]

Test 1 (x)	15	23	25	30	34	34	40
Test 2 (y)	20	26	27	32	35	37	35

(a) Find the value of a and the value of b . [2]

(b) Find the value of p and the value of q . [3]

[illegible]

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 15]

The length, X mm, of a certain species of seashell is normally distributed with mean 25 and variance, σ^2 .

The probability that X is less than 24.15 is 0.1446.

(a) Find $P(24.15 < X < 25)$. [2]

(b) (i) Find σ , the standard deviation of X .

(ii) Hence, find the probability that a seashell selected at random has a length greater than 26 mm. [5]

A random sample of 10 seashells is collected on a beach. Let Y represent the number of seashells with lengths greater than 26 mm.

(c) Find $E(Y)$. [3]

(d) Find the probability that exactly three of these seashells have a length greater than 26 mm. [2]

A seashell selected at random has a length less than 26 mm.

(e) Find the probability that its length is between 24.15 mm and 25 mm. [3]



Turn over

— 12 —

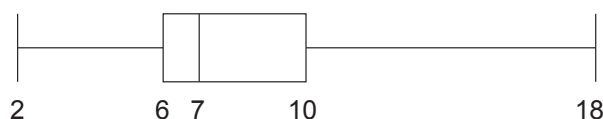
0120/51/MAT/AR/11/2/ENG/120/XX

Do **not** write solutions on this page.

2. [Maximum mark: 7]

The ages of the eldest child are summarized in the following box and whisker diagram.

ages of eldest child (years)



- (a) Find the largest value of c that would not be considered an outlier. [3]

The regression line of a on c is $a = \frac{7}{4}c + 20$. The regression line of c on a is $c = \frac{1}{2}a - 9$.

- (b) (i) One of the adults surveyed is 42 years old. Estimate the age of their eldest child.

- (ii) Find the mean age of all the adults surveyed. [4]

[illegible]

Turn over

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 16]

A biased four-sided die with faces labelled 1, 2, 3 and 4 is rolled and the result recorded. Let X be the result obtained when the die is rolled. The probability distribution for X is given in the following table where p and q are constants.

x	1	2	3	4
$P(X = x)$	p	0.3	q	0.1

For this probability distribution, it is known that $E(X) = 2$.

(a) Show that $p = 0.4$ and $q = 0.2$. [5]

(b) Find $P(X > 2)$. [2]

Nicky plays a game with this four-sided die. In this game she is allowed a maximum of five rolls. Her score is calculated by adding the results of each roll. Nicky wins the game if her score is at least ten.

After three rolls of the die, Nicky has a score of four.

(c) Assuming that rolls of the die are independent, find the probability that Nicky wins the game. [5]

David has two pairs of unbiased four-sided dice, a yellow pair and a red pair. Both yellow dice have faces labelled 1, 2, 3 and 4. Let S represent the sum obtained by rolling the two yellow dice. The probability distribution for S is shown below.

s	2	3	4	5	6	7	8
$P(S = s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

The first red die has faces labelled 1, 2, 2 and 3. The second red die has faces labelled 1, a , a and b , where $a < b$ and $a, b \in \mathbb{Z}^+$. The probability distribution for the sum obtained by rolling the red pair is the same as the distribution for the sum obtained by rolling the yellow pair.

(d) Determine the value of b . [2]

(e) Find the value of a , providing evidence for your answer. [2]



— 10 —

44447111

Do **not** write solutions on this page.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

1. [Maximum mark: 4]

The number of hours spent exercising each week by a group of students is shown in the following table.

Exercising time (in hours)	Number of students
2	5
3	1
4	4
5	3
6	x

The median is 4.5 hours.

- (a) Find the value of x .

[2]
- (b) Find the standard deviation.

[2]

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

11. [Maximum mark: 16]

A bakery makes two types of muffins: chocolate muffins and banana muffins.

The weights, C grams, of the chocolate muffins are normally distributed with a mean of 62 g and standard deviation of 2.9 g.

(a) Find the probability that a randomly selected chocolate muffin weighs less than 61 g. [2]

(b) In a random selection of 12 chocolate muffins, find the probability that exactly 5 weigh less than 61 g. [2]

The weights, B grams, of the banana muffins are normally distributed with a mean of 68 g and standard deviation of 3.4 g.

Each day 60% of the muffins made are chocolate.

On a particular day, a muffin is randomly selected from all those made at the bakery.

(c) (i) Find the probability that the randomly selected muffin weighs less than 61 g.
(ii) Given that a randomly selected muffin weighs less than 61 g, find the probability that it is chocolate. [7]

The machine that makes the chocolate muffins is adjusted so that the mean weight of the chocolate muffins remains the same but their standard deviation changes to σ g. The machine that makes the banana muffins is not adjusted. The probability that the weight of a randomly selected muffin from these machines is less than 61 g is now 0.157.

(d) Find the value of σ . [5]



Do **not** write solutions on this page

Section A

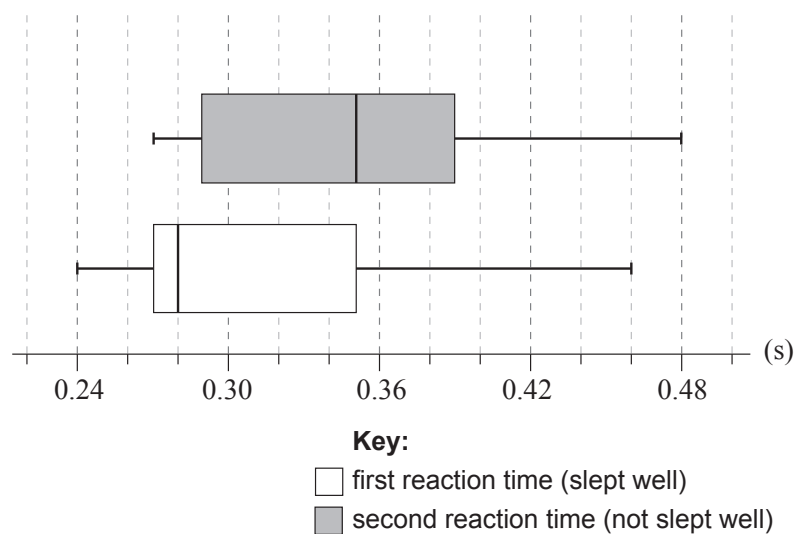
Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

4. [Maximum mark: 6]

A random sample of nine adults were selected to see whether sleeping well affected their reaction times to a visual stimulus. Each adult's reaction time was measured twice.

The first measurement for reaction time was taken on a morning after the adult had slept well. The second measurement was taken on a morning after the same adult had not slept well.

The box and whisker diagrams for the reaction times, measured in seconds, are shown below.



Consider the box and whisker diagram representing the reaction times after sleeping well.

- (a) State the median reaction time after sleeping well. [1]
- (b) Verify that the measurement of 0.46 seconds is not an outlier. [3]
- (c) State why it appears that the mean reaction time is greater than the median reaction time. [1]

Now consider the two box and whisker diagrams.

- (d) Comment on whether these box and whisker diagrams provide any evidence that might suggest that not sleeping well causes an increase in reaction time. [1]

(This question continues on the following page)



[illegible]

Apollo | Original source crop | 39

6. [Maximum mark: 6]

(c) Determine the maximum possible value of $P(A \cap B)$, justifying your answer. [2]

[illegible]

1. [Maximum mark: 5]

Mathematics scores (x)	64	68	72	75	80	82	85	86
Science scores (y)	67	72	77	76	84	83	89	91

(a) Find the value of a and the value of b . [2]

(b) Write down the value of the Pearson's product-moment correlation coefficient, r . [1]

(c) Use the equation of your regression line to predict the Science test score for a student who has a score of 78 on the Mathematics test. Express your answer to the nearest integer. [2]



Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 16]

The time worked, T , in hours per week by employees of a large company is normally distributed with a mean of 42 and standard deviation 10.7.

- (a) Find the probability that an employee selected at random works more than 40 hours per week. [2]

- (b) A group of four employees is selected at random. Each employee is asked in turn whether they work more than 40 hours per week. Find the probability that the fourth employee is the only one in the group who works more than 40 hours per week. [3]

- (c) A large group of employees work more than 40 hours per week.

- (i) An employee is selected at random from this large group.

Find the probability that this employee works less than 55 hours per week.

- (ii) Ten employees are selected at random from this large group.

Find the probability that exactly five of them work less than 55 hours per week. [7]

It is known that $P(a \leq T \leq b) = 0.904$ and that $P(T > b) = 2P(T < a)$, where a and b are numbers of hours worked per week. An employee who works fewer than a hours per week is considered to be a part-time employee.

- (d) Find the maximum time, in hours per week, that an employee can work and still be considered part-time. [4]



Turn over

Do **not** write solutions on this page.

2. [Maximum mark: 6]

Number of times on <i>The Dragon</i>	Frequency
0	6
1	16
2	13
3	2
4	3

[4]

[2]

[illegible]

6. [Maximum mark: 5]

$$f(x) = \begin{cases} \frac{1}{2a}, & a \leq x \leq 3a \\ 0, & \text{otherwise} \end{cases}$$

(a) State $E(X)$ in terms of a . [1]

(b) Use integration to find $\text{Var}(X)$ in terms of a . [4]

[illegible]

Apollo | Original source crop | 48

Section A

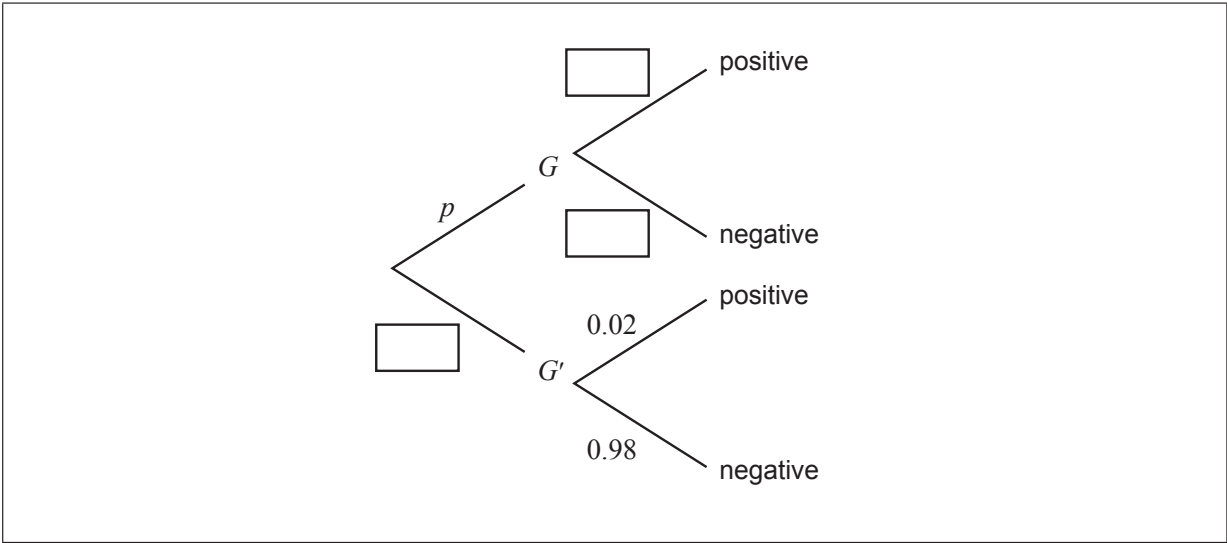
Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

7. [Maximum mark: 6]

A new test has been developed to identify whether a particular gene, G , is present in a population of parrots. The test returns a correct positive result 95% of the time for parrots with the gene, and a false positive result 2% of the time for parrots without the gene.

The proportion of the population with the gene is p .

(a) Complete the tree diagram below. [2]



(b) A random sample of the population was tested. It was found that 150 tests returned a positive result. Out of the 150 parrots with a positive test result, 18 did not actually have the gene. Find an estimate for p . [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

11. [Maximum mark: 19]

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

$$f(x) = \begin{cases} \frac{6}{\pi\sqrt{16-x^2}}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the exact value of $E(X)$. [5]

(b) Find $P(X < 0.5)$. [2]

A laboratory trial may require up to 2 millilitres of reagent. The amount of reagent used has been found to have a probability distribution that can be modelled by $f(x)$, where X is the amount of reagent in millilitres.

Each laboratory trial is independent. A trial is considered a success when $X < 0.5$.

(c) Determine the least number of trials required to be 99% sure of at least one success. [3]

Ten trials were conducted.

(d) Find the probability that exactly three trials were successful. [2]

(e) Write down the number of ways these three successful trials could have occurred consecutively. [1]

Now consider n trials where it is given that exactly three successes have occurred.

(f) (i) Write down an expression for the number of ways these three successful trials could have occurred consecutively.
(ii) Find the greatest value of n such that the probability of three consecutive successful trials is more than 0.05. [6]



Do **not** write solutions on this page.



Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

11. [Maximum mark: 20]

A game of chance involves drawing **two** balls at random out of a box without replacement. The box initially contains r red balls and y yellow balls.

Let $P(YY)$ represent the probability of drawing two yellow balls from the box without replacement.

Consider a version of this game where it is known that $P(YY) = \frac{1}{3}$.

(a) Show that $2y^2 - 2(r+1)y + r - r^2 = 0$. [4]

(b) By solving the equation in part (a), show that $y = \frac{(r+1) + \sqrt{3r^2+1}}{2}$. [4]

(c) Find two pairs of values for r and y that satisfy the condition $P(YY) = \frac{1}{3}$. [4]

Now consider a similar game of chance that involves drawing **three** balls out of a box without replacement. The box initially contains 10 red balls and y yellow balls.

Let $P(YYY)$ represent the probability of drawing three yellow balls from the box without replacement.

(d) Find an expression for $P(YYY)$ in terms of y . [3]

A yellow ball is added so that the box now contains 10 red balls and $(y+1)$ yellow balls. The probability of drawing three yellow balls from the box without replacement is now twice the probability expressed in part (d).

(e) Find the initial number of yellow balls in the box. [5]



Do **not** write solutions on this page.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

2. [Maximum mark: 4]

Events A and B are such that $P(A) = 0.65$, $P(B) = 0.75$ and $P(A \cap B) = 0.6$.

(a) Find $P(A \cup B)$. [2]

(b) Hence, or otherwise, find $P(A' \cap B')$. [2]

This image shows a full page of white paper with horizontal dotted lines, typical of primary school handwriting practice paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

5. [Maximum mark: 5]

x	1	2	3	4
$P(X=x)$	k	k^2	a	k^3

[illegible]

6. [Maximum mark: 5]

(a) Find the possible values of p . [3]

(b) Find $\text{Var}(Y)$. [2]

This image shows a full page of white paper with horizontal dotted lines, typical of primary school writing paper. The lines are evenly spaced and run across the entire width of the page. There are no margins, text, or other markings present.

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 16]

A farmer is growing a field of wheat plants. The height, H cm, of each plant can be modelled by a normal distribution with mean μ and standard deviation σ .

It is known that $P(H < 94.6) = 0.288$ and $P(H > 98.1) = 0.434$.

- (a) Find the probability that the height of a randomly selected plant is between 94.6 cm and 98.1 cm. [2]
- (b) Find the value of μ and the value of σ . [5]

The farmer measures 100 randomly selected plants. Any plant with a height greater than 98.1 cm is considered ready to harvest. Heights of plants are independent of each other.

- (c) (i) Find the probability that exactly 34 plants are ready to harvest.
- (ii) Given that fewer than 49 plants are ready to harvest, find the probability that exactly 34 plants are ready to harvest. [6]

In another field, the farmer is growing the same variety of wheat, but is using a different fertilizer. The heights of these plants, F cm, are normally distributed with mean 98.6 and standard deviation d . The farmer finds the interquartile range to be 4.82 cm.

- (d) Find the value of d . [3]



Do **not** write solutions on this page.

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 16]

A farmer is growing a field of wheat plants. The height, H cm, of each plant can be modelled by a normal distribution with mean μ and standard deviation σ .

It is known that $P(H < 94.6) = 0.288$ and $P(H > 98.1) = 0.434$.

(a) Find the probability that the height of a randomly selected plant is between 94.6 cm and 98.1 cm. [2]

(b) Find the value of μ and the value of σ . [5]

The farmer measures 100 randomly selected plants. Any plant with a height greater than 98.1 cm is considered ready to harvest. Heights of plants are independent of each other.

(c) (i) Find the probability that exactly 34 plants are ready to harvest.
(ii) Given that fewer than 49 plants are ready to harvest, find the probability that exactly 34 plants are ready to harvest. [6]

In another field, the farmer is growing the same variety of wheat, but is using a different fertilizer. The heights of these plants, F cm, are normally distributed with mean 98.6 and standard deviation d . The farmer finds the interquartile range to be 4.82 cm.

(d) Find the value of d . [3]



Do **not** write solutions on this page.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

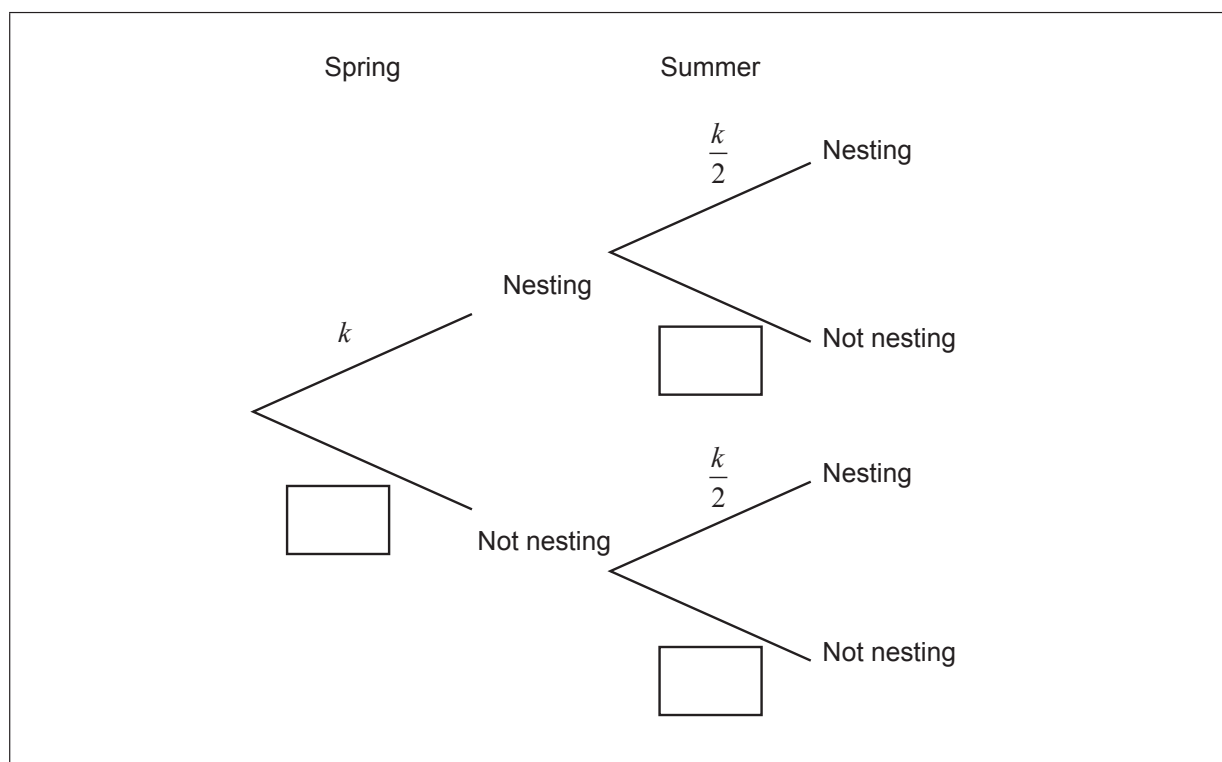
4. [Maximum mark: 6]

A species of bird can nest in two seasons: Spring and Summer.

The probability of nesting in Spring is k .

The probability of nesting in Summer is $\frac{k}{2}$.

This is shown in the following tree diagram.



- (a) Complete the tree diagram to show the probabilities of not nesting in each season. Write your answers in terms of k .

[2]

It is known that the probability of not nesting in Spring and not nesting in Summer is $\frac{5}{9}$.

- (b) (i) Show that $9k^2 - 27k + 8 = 0$.
- (ii) Both $k = \frac{1}{3}$ and $k = \frac{8}{3}$ satisfy $9k^2 - 27k + 8 = 0$.

State why $k = \frac{1}{3}$ is the only valid solution.

[4]

(This question continues on the following page)



[illegible]

Apollo | Original source crop | 70

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

5. [Maximum mark: 7]

A class is given two tests, Test A and Test B. Each test is scored out of a total of 100 marks. The scores of the students are shown in the following table.

Student	1	2	3	4	5	6	7	8	9	10
Test A	52	71	100	93	81	80	88	100	70	61
Test B	58	80	92	98	90	82	100	100	65	74

Let x be the score on Test A and y be the score on Test B.

The teacher finds that the equation of the regression line of y on x for these scores is $y = 0.822x + 18.4$.

(a) Find the value of Pearson’s product-moment correlation coefficient, r . [2]

Giovanni was absent for Test A and Paulo was absent for Test B.

The teacher uses the regression line of y on x to estimate the missing scores.

Paulo scored 10 on Test A.

The teacher estimated his score on Test B to be 27 to the nearest integer using the following calculation:

$$y = 0.822(10) + 18.4 \approx 27$$

(b) Give a reason why this method is not appropriate for Paulo. [1]

Giovanni scored 90 on Test B.

The teacher estimated his score on Test A to be 87 to the nearest integer using the following calculation:

$$90 = 0.822x + 18.4, \text{ so } x = \frac{90 - 18.4}{0.822} \approx 87$$

(c) (i) Give a reason why this method is not appropriate for Giovanni.
(ii) Use an appropriate method to show that the estimated Test A score for Giovanni is 86 to the nearest integer. [4]

(This question continues on the following page)



[illegible]

Apollo | Original source crop | 74

2. [Maximum mark: 5]

x	5	6	6	8	10
y	9	13	p	q	21

The regression line passes through the mean point $(\bar{x}, \bar{y})$.

- (a) Given that $\bar{x} = 7$, verify that $\bar{y} = 16$. [1]
- (b) Given that $q - p = 3$, find the value of p and the value of q . [4]

[illegible]

Apollo | Original source crop | 77

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 15]

A shop sells chocolates. The weight, in kilograms, of chocolates bought by a random customer can be modelled by a continuous random variable X with probability density function f defined by

$$f(x) = \begin{cases} \frac{6}{85}(4 + 3x - x^2), & 0.5 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}.$$

- (a) Find the mode of X . [2]
- (b) Find $P(1 \leq X \leq 2)$. [2]
- (c) Find the median of X . [3]

The shop sells chocolates to customers at \$25 per kilogram.

However, if the weight of chocolate bought by a customer is at least 0.75 kilograms, the shop sells chocolate at a discounted rate of \$24 per kilogram.

- (d) Find the probability that a randomly selected customer spends at most \$48. [3]
- (e) Find the expected amount spent per customer. Give your answer correct to the nearest cent. [5]



Turn over

Do **not** write solutions on this page.

1. [Maximum mark: 27]

This question considers two possible models for the occurrence of random events in a computer game.

In a new computer game, each time a player performs an action, there is a random chance that the action will be *boosted*, meaning that it provides a benefit to the player.

The designer of this computer game is considering two possible models for when to boost an action.

In the first model, the probability that an action will be boosted is constant.

(a) Suppose the probability that an action will be boosted is 0.1.

(i) Find the probability that the first boost occurs on the third action. [2]

(ii) Find the probability that at least one boost occurs in the first six actions. [3]

(b) Suppose the probability that an action will be boosted is p , where $0 < p < 1$.

(i) Explain why the probability that the first boost occurs on the x^{th} action is $p(1-p)^{x-1}$. [1]

Let X be the number of actions until the first boost occurs.

(ii) Hence, write down an expression, using sigma notation, for $E(X)$ in terms of x and p . [1]

Consider the sum of an infinite geometric sequence, with first term a and common ratio r ($|r| < 1$),

$$a + ar + ar^2 + ar^3 + \dots = \frac{a}{1-r}.$$

(c) (i) By differentiating both sides of the above equation with respect to r , find an expression for $\sum_{n=1}^{\infty} nar^{n-1}$ in terms of a and r . [4]

(ii) Hence, show that $E(X) = \frac{1}{p}$. [2]

(This question continues on the following page)

(Question 1 continued)

It can be shown that $\text{Var}(X) = \frac{1-p}{p^2}$.

- (d) Find $E(X)$ and $\text{Var}(X)$ when $p = 0.1$. [2]

In the designer’s second model, the initial probability that an action is boosted is 0.2, and each time an action occurs that is not boosted, the probability that the next action is boosted increases by 0.2. After an action has been boosted, the probability resets to 0.2 for the next action.

- (e) Show that the probability that the first boost occurs on the third action is 0.288. [2]

Let Y be the number of actions until the first boost occurs.

- (f) Explain why $Y \leq 5$. [1]

The following table shows the probability distribution of Y .

y	1	2	3	4	5
$P(Y = y)$	0.2	m	0.288	n	0.0384

- (g) (i) Find the value of m and the value of n . [2]
- (ii) Show that $E(Y) = 2.5104$. [2]
- (iii) Find $\text{Var}(Y)$. [2]
- (h) (i) Use the expression given in (c)(ii) to find the value of p for which $E(X) = E(Y)$. [1]
- (ii) Find $\text{Var}(X)$ for this value of p . [1]
- (iii) Hence determine, with a reason, which model provides a more consistent experience for the player with respect to boosted actions. [1]

2. [Maximum mark: 31]

This question asks you to find the probability of graphs of randomly generated quadratic functions having a specified number of x -intercepts.

In parts (a) – (f), consider quadratic functions, $f(x) = ax^2 + bx + c$, whose coefficients, a , b and c , are randomly generated in turn by rolling an unbiased six-sided die three times and reading off the value shown on the uppermost face of the die.

For example, rolling a 2, 3 and 5 in turn generates the quadratic function $f(x) = 2x^2 + 3x + 5$.

- (a) Explain why there are 216 possible quadratic functions that can be generated using this method. [1]

- (b) The set of coefficients, $a = 1$, $b = 4$ and $c = 4$, is randomly generated to form the quadratic function $f(x) = x^2 + 4x + 4$.

Verify that this graph of f has only one x -intercept. [2]

- (c) By considering the discriminant, or otherwise, show that the probability of the graph of such a randomly generated quadratic function having only one x -intercept is $\frac{5}{216}$. [6]

Now consider randomly generated quadratic functions whose corresponding graphs have two **distinct** x -intercepts.

- (d) By considering the discriminant, determine the set of possible values of ac . [3]

- (e) (i) For the case where $ac = 1$, show that there are four quadratic functions whose corresponding graphs have two distinct x -intercepts. [1]

- (ii) For the case where $ac = 2$, show that there are eight quadratic functions whose corresponding graphs have two distinct x -intercepts. [2]

Let p be the probability of the graph of such a randomly generated quadratic function having two distinct x -intercepts.

- (f) Using the approach started in part (e), or otherwise, find the value of p . [6]

(This question continues on the following page)

(Question 2 continued)

In parts (g) and (h), consider a randomly generated quadratic function, $f(x) = x^2 + 2Zx + 1$, where the continuous random variable $Z \sim N(0, 1)$.

- (g) Find the probability that the graph of f has two x -intercepts. [3]

The continuous random variables, X_1 and X_2 , represent the x -intercepts of the graph of f where $X_1 = -Z - \sqrt{Z^2 - 1}$ and $X_2 = -Z + \sqrt{Z^2 - 1}$.

- (h) Given that the graph of f has two x -intercepts, X_1 and X_2 , find the probability that both X_1 and X_2 are greater than 0.5. [7]

3. [Maximum mark: 5]

$$P(X=x) = \frac{kx}{20} \text{ for } x \in \{3, 5, 8, 11\}.$$

- (a) Find the value of k . [2]
- (b) Find $E(X)$. [3]

[illegible]

Apollo | Original source crop | 86

6. [Maximum mark: 8]

$$f(x) = \begin{cases} \frac{1}{5} & 0 \leq x < 2 \\ -\frac{x}{30} + \frac{4}{15} & 2 \leq x \leq 8 \\ 0 & \text{otherwise.} \end{cases}$$

- Find $E(X)$. [3]
- Given that $E(c - 2X) = 0$, where c is a constant, determine the value of c . [2]
- Find the median of X . [3]

[illegible]

4. [Maximum mark: 8]

(c) Find the value of σ . [3]

[illegible]

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

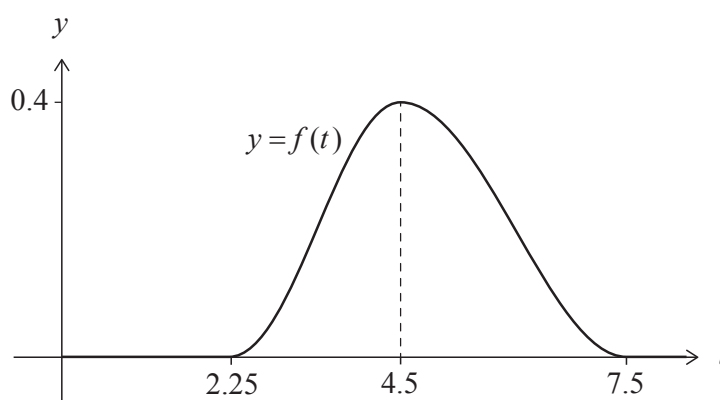
11. [Maximum mark: 19]

In a marathon race, the random variable T represents the time, in hours, taken for a runner to complete the race. No runner completes the race in less than 2.25 hours, and no runner completes it in more than 7.5 hours.

The probability density function for T is modelled by f , defined by

$$f(t) = \begin{cases} \frac{4}{21} \left(1 - \cos \left(\frac{4\pi}{9} (t - 2.25) \right) \right), & 2.25 \leq t < 4.5 \\ \frac{4}{21} \left(1 + \cos \left(\frac{\pi}{3} (t - 4.5) \right) \right), & 4.5 \leq t \leq 7.5 \\ 0, & \text{otherwise.} \end{cases}$$

The graph of f has a maximum point at $t = 4.5$ as shown in the following diagram:



- (a) (i) Find the value of $\int_{2.25}^{4.5} f(t) \, dt$.
- (ii) Write down the mode of T .
- (iii) Determine which is greater, the mode of T or the median of T , justifying your answer.

[4]

(This question continues on the following page)



Do **not** write solutions on this page.

(Question 11 continued)

The runners who finish the race in 3.5 hours or less are considered to be fast runners.

- (b) Find the probability that a runner chosen at random is a fast runner. [2]
- (c) Find the probability that a fast runner chosen at random finishes the race in 3 hours or less. [3]
- (d) Find the lower quartile of T . [3]

Each runner's time is converted to a score which is calculated as $a - bt$, where t represents their time in hours, and $a, b > 0$.

Consider the random variable P which represents the score of a runner. It is given that $E(P) = 100$ and the maximum possible score is 150.

- (e) Use $E(T) = 4.723$ to determine the value of a and the value of b , giving your answers to the nearest integer. [5]
- (f) Given also that $\text{Var}(T) = 0.906$, find $\text{Var}(P)$. [2]



Turn over

Do **not** write solutions on this page.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

1. [Maximum mark: 4]

The following table shows the number of hours of play time, x , and sleep time, y , for a group of six children, over the period of one week.

Play time (x)	11	13	14	17	22	24
Sleep time (y)	62	65	68	75	84	87

The regression line of y on x for this data can be written in the form $y = ax + b$.

- (a) Find the value of a and the value of b . [2]
- (b) Use the equation of the regression line to estimate the sleep time of a child whose weekly play time is 20 hours. [2]

[illegible]

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

10. [Maximum mark: 17]

At Adam's Apple Orchard the weights of apples, W , in grams, are normally distributed with a mean 175 grams and standard deviation 8 grams.

- (a) Find the probability that a randomly chosen apple weighs less than 170 grams. [2]
- (b) It is found that 20% of the apples weigh more than w grams. Find w , correct to four significant figures. [2]

All orchards classify an apple as premium when its weight is between 170 and 185 grams.

- (c) Find the percentage of apples that are classified as premium at Adam's Apple Orchard. [2]

After orders are completed, there are many apples left over. Boxes are filled with randomly chosen left-over apples. Each box contains 40 apples.

- (d) Find the probability that a randomly chosen box contains at least 30 premium apples. [3]
- (e) If 10 of these boxes are randomly selected, find the probability that exactly 4 boxes have at least 30 premium apples. [2]

At a neighbouring orchard the weights of apples, M , in grams, are normally distributed with mean μ and standard deviation σ . It is known that:

- 82% of their apples are classified as premium
- the percentage of apples that weigh less than 170 grams is twice the percentage of apples that weigh more than 185 grams.

- (f) Find the value of μ . [6]



Turn over

— 16 —

444571116

Do **not** write solutions on this page.

Section A

Answer **all** questions. Answers must be written within the answer boxes provided. Working may be continued below the lines, if necessary.

3. [Maximum mark: 6]

A supermarket analyses the shopping habits of its customers.

The number of times, X , each customer visits the supermarket in a week is given by the following probability distribution.

x	1	2	3	4	5	≥ 6
$P(X = x)$	$1.5a$	$2a$	0.281	a	0.026	0

- (a)

(i) Find the value of a .

(ii) Write down the mode of X .

[3]
- (b)

Find the mean of X .

[2]

The manager wants to know why customers come to their supermarket. They survey the first 50 customers to arrive at the supermarket on a particular day.

- (c)

Identify which one of the following best describes the manager's sampling method.
Circle your answer.

[1]

Simple random / Systematic / Convenience / Quota / Stratified

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

11. [Maximum mark: 18]

Amanda enters data from surveys into a database. It can be assumed that the accuracy of any survey entered is independent of all other surveys entered.

From previous records, it is known that Amanda enters 8% of the surveys inaccurately.

- (a) On a particular day Amanda enters data from 50 surveys.
- (i) Find the probability that Amanda entered at most six surveys inaccurately.
 - (ii) Given that at most six surveys were entered inaccurately, find the probability that exactly four surveys were entered inaccurately. [5]

On a different day Amanda enters data from n surveys. On this day, the probability that at most six surveys were entered inaccurately is approximately 0.367.

- (b) Find the value of n . [3]

Bryce and Carmen also enter data from surveys into the same database. It is known that surveys entered by Bryce and Carmen are inaccurate 6% and 11% of the time respectively. It can again be assumed that the accuracy of any survey entered is independent of all other surveys entered.

From the surveys assigned to the three of them, Amanda enters 55%, Bryce 25% and Carmen 20%.

- (c) Find the probability that a randomly selected survey was
- (i) entered inaccurately;
 - (ii) entered by Amanda, given that the survey was entered inaccurately. [6]

The following year, the accuracy of Amanda's and Bryce's work remained the same, as did the percentage of surveys entered by each of the three employees. However, Carmen's accuracy had improved and the probability that she entered a survey inaccurately was now $x\%$.

The probability that a randomly selected survey had been entered inaccurately was now the same as the probability that Carmen made an error when entering a survey.

- (d) Find the value of x . [4]



Do **not** write solutions on this page.

1. [Maximum mark: 23]

This question asks you to use polynomial functions to model some situations in probability.

Two unbiased tetrahedral (four-sided) dice with faces labelled 1, 2, 3 and 4 are thrown and the scores recorded.

The random variable M denotes the maximum of these two scores.

The probability distribution of M is given in the following table.

m	1	2	3	4
$P(M = m)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$

(a) Find $E(M)$.

[2]

An alternative way to represent the probability distribution of M is to use a polynomial

function, G , where $G(t) = \sum_{m=1}^4 P(M = m)t^m$.

Hence, for the distribution of M , $G(t) = \frac{1}{16}t + \frac{3}{16}t^2 + \frac{5}{16}t^3 + \frac{7}{16}t^4$.

(b) Find $G(1)$.

[1]

(c) (i) Find $G'(t)$.

[2]

(ii) Hence, show that $G'(1) = E(M)$.

[3]

(This question continues on the following page)

(Question 1 continued)

A bag contains two red balls and three yellow balls.

Two balls are selected at random without replacement from the bag.

The random variable X denotes the total number of red balls selected.

The probability distribution of X can be represented by the polynomial function, G_X , where

$$G_X(t) = \sum_{x=0}^2 P(X=x)t^x.$$

- (d) Show that $G_X(t) = \frac{3}{10} + \frac{3}{5}t + \frac{1}{10}t^2$, making it clear how the coefficients of $G_X(t)$ have been determined. [5]

An unbiased coin and a biased coin are tossed.

The probability of obtaining a tail on the biased coin is p .

The random variable Y denotes the total number of tails obtained from tossing both coins.

The probability distribution of Y can be represented by the polynomial function, G_Y , where

$$G_Y(t) = \sum_{y=0}^2 P(Y=y)t^y.$$

- (e) Given that the coefficient of t^2 in $G_Y(t)$ is $\frac{1}{3}$, find

(i) the value of p ; [2]

(ii) an expression for $G_Y(t)$. [4]

The random variable Z denotes the sum of the total number of red balls selected, X , and the total number of tails obtained from tossing both coins, Y .

The probability distribution of Z can be represented by the function, G_Z , where

$$G_Z(t) = G_X(t)G_Y(t).$$

- (f) For random variable Z , it can be shown that $G_Z'(1) = E(Z)$.

Use this result to find $E(Z)$. [4]

Turn over

3. [Maximum mark: 7]

x	3.8	6.3	2.5	2.4	8.9	10.7	6.7	10.3
y	11.8	9.7	6.1	4.1	p	23.3	13.7	24.5

- (a) Find $\bar{x}$. [2]

(b) Find

- (i) $\bar{y}$;
- (ii) the value of p . [4]

- (c) Hence, find Pearson's product-moment correlation coefficient, r . [1]

[illegible]

5. [Maximum mark: 5]

Class	Frequency
$15 < y \leq 20$	31
$20 < y \leq 25$	42
$25 < y \leq 30$	61
$30 < y \leq 35$	46
$35 < y \leq 40$	29

- (ii) the standard deviation.

[2]

(b) Use this information and your mean and standard deviation found in part (a) to find an approximate value for the interquartile range correct to two significant figures.

[3]

This image shows a full page of white paper with horizontal dotted lines, typical of primary school writing paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

7. [Maximum mark: 7]

Both machines are equally likely to be selected.

- (a) Find the probability that exactly two semiconductors are defective. [4]
- (b) Given that exactly two semiconductors are defective, find the probability that they were manufactured by machine A. [3]

[illegible]

Apollo | Original source crop | 108

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

11. [Maximum mark: 18]

- (a) Use integration by parts to find $\int \arccos x \, dx$. [4]

The probability density function of a random variable X , is given by

$$f(x) = \begin{cases} 3x \arccos(x^2), & 0 \leq x \leq k; \\ 0, & \text{otherwise} \end{cases}.$$

- (b) (i) Show that $k^2 \arccos(k^2) - \sqrt{1-k^4} + \frac{1}{3} = 0$.
 (ii) Hence, find the value of k . Give your answer correct to six significant figures. [6]
- (c) Find
 (i) $E(X)$;
 (ii) $\text{Var}(X)$. [5]
- (d) Given that $\mu = E(X)$ and $\sigma^2 = \text{Var}(X)$, find $P(\mu - \sigma < X < \mu + \sigma)$. [3]



— 11 —

0020-7107

Do **not** write solutions on this page.

2. [Maximum mark: 6]

Aiden and Brett were absent on the day of the quiz.

A box plot showing the distribution of quiz scores. The x-axis is labeled 'quiz score' and ranges from 6 to 17. The box plot has a minimum at 6, a first quartile (Q1) at approximately 6.5, a median at approximately 10.5, a third quartile (Q3) at approximately 14.5, and a maximum at 17.

Brett scores more than 17.

(a) Explain, briefly, why the median score for all 30 pupils would still be 10.5. [1]

The mean score for all 30 pupils is now 10.6.

The range of scores for all 30 pupils is 14.

(b) Determine Aiden's score and Brett's score. [5]

[illegible]

Section B

Answer **all** questions in the answer booklet provided. Please start each question on a new page.

11. [Maximum mark: 17]

The lengths, in metres, of jumps in a long jump competition can be modelled by a continuous random variable X with probability density function f defined by:

$$f(x) = \begin{cases} 0, & x < 3 \\ a(x-3)^3 + b(x-3)^2, & 3 \leq x \leq 9 \\ 0, & x > 9 \end{cases}$$

where $a, b \in \mathbb{R}$, $a, b \neq 0$.

- (a) Show that $324a + 72b = 1$. [4]

It is given that $f(9) = 0$.

- (b) (i) Show that $6a + b = 0$.
 (ii) Determine the value of a and the value of b . [4]
- (c) Show that the median of X is less than the mode of X . [5]

Matt has the final jump in the competition. He needs to jump at least 8.52 m to win the competition.

- (d) Given that he jumps over 8 m, use the model to find the probability that he wins the competition. [4]



Do **not** write solutions on this page.