Subject – Math AA(Standard Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2024
Paper -2
Questions

Question 1

[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]

Question 2

[Maximum mark: 5]

The following table below shows the marks scored by seven students on two different mathematics tests.

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

Let L_1 be the regression line of x on y. The equation of the line L_1 can be written in the form $x = ay + b$.

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

Let L_2 be the regression line of y on x. The lines L_1 and L_2 pass through the same point with coordinates (p, q).

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

Question 3

[Maximum mark: 6]

A six-sided biased die is weighted in such a way that the probability of obtaining a "six" is $\frac{7}{10}$.

The die is tossed five times. Find the probability of obtaining

(a) at most three "sixes". [3]

(b) the third "six" on the fifth toss. [3]

Question 4

[Maximum mark: 6]

A company produces bags of sugar whose masses, in grams, can be modelled by a normal distribution with mean 1000 and standard deviation 3.5. A bag of sugar is rejected for sale if its mass is less than 995 grams.

(a) Find the probability that a bag selected at random is rejected. [2]

(b) Estimate the number of bags which will be rejected from a random sample of 100 bags. [1]

(c) Given that a bag is not rejected, find the probability that it has a mass greater than 1005 grams. [3]

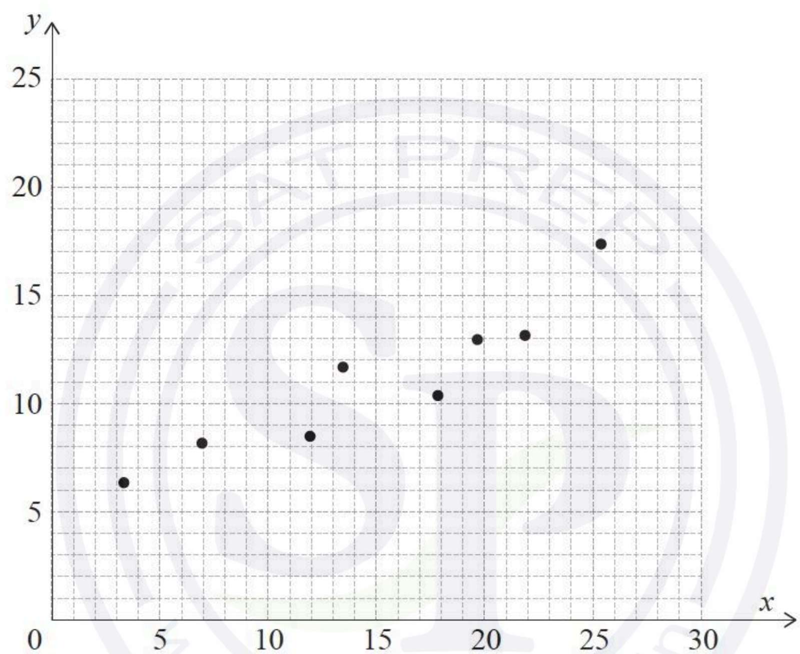
Question 5

[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.



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]

Question 6

[Maximum mark: 15]

All answers in this question should be given to four significant figures.

In a local weekly lottery, tickets cost \$2 each.

In the first week of the lottery, a player will receive $\$D$ for each ticket, with the probability distribution shown in the following table. For example, the probability of a player receiving \$10 is 0.03. The grand prize in the first week of the lottery is \$1000.

d	0	2	10	50	Grand Prize
$P(D=d)$	0.85	c	0.03	0.002	0.0001

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

(b) Determine whether this lottery is a fair game in the first week. Justify your answer. [4]

If nobody wins the grand prize in the first week, the probabilities will remain the same, but the value of the grand prize will be \$2000 in the second week, and the value of the grand prize will continue to double each week until it is won. All other prize amounts will remain the same.

(c) Given that the grand prize is not won and the grand prize continues to double, write an expression in terms of n for the value of the grand prize in the nth week of the lottery. [2]

The wth week is the first week in which the player is expected to make a profit. Ryan knows that if he buys a lottery ticket in the wth week, his expected profit is $\$p$.

(d) Find the value of p. [7]

Question 7

[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]

Question 8

[Maximum mark: 8]

At a school, 70% of the students play a sport and 20% of the students are involved in theatre. 18% of the students do neither activity.

A student is selected at random.

(a) Find the probability that the student plays a sport and is involved in theatre. [2]

(b) Find the probability that the student is involved in theatre, but does not play a sport. [2]

At the school 48% of the students are girls, and 25% of the girls are involved in theatre.

A student is selected at random. Let G be the event "the student is a girl" and let T be the event "the student is involved in theatre".

(c) Find $\mathrm{P}(G \cap T)$. [2]

(d) Determine if the events G and T are independent. Justify your answer. [2]

Question 9

[Maximum mark: 6]

At a café, the waiting time between ordering and receiving a cup of coffee is dependent upon the number of customers who have already ordered their coffee and are waiting to receive it.

Sarah, a regular customer, visited the café on five consecutive days. The following table shows the number of customers, x, ahead of Sarah who have already ordered and are waiting to receive their coffee and Sarah's waiting time, y minutes.

Number of customers (x)	3	9	11	10	5
Sarah's waiting time (y)	6	10	12	11	6

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a) (i) Find the value of a and the value of b.

(ii) Write down the value of Pearson's product-moment correlation coefficient, r. [3]

(b) Interpret, in context, the value of a found in part (a)(i). [1]

On another day, Sarah visits the café to order a coffee. Seven customers have already ordered their coffee and are waiting to receive it.

(c) Use the result from part (a)(i) to estimate Sarah's waiting time to receive her coffee. [2]

Question 10

[Maximum mark: 16]

The random variable X follows a normal distribution with mean μ and standard deviation σ.

(a) Find $P(\mu - 1.5\sigma < X < \mu + 1.5\sigma)$. [3]

The avocados grown on a farm have weights, in grams, that are normally distributed with mean μ and standard deviation σ. Avocados are categorized as small, medium, large or premium, according to their weight. The following table shows the probability an avocado grown on the farm is classified as small, medium, large or premium.

Category	Small	Medium	Large	Premium
Probability	0.04	0.576	0.288	0.096

The maximum weight of a small avocado is 106.2 grams.

The minimum weight of a premium avocado is 182.6 grams.

(b) Find the value of μ and of σ. [5]

A supermarket purchases all the avocados from the farm that weigh more than 106.2 grams.

(c) Find the probability that an avocado chosen at random from this purchase is categorized as

(i) medium;

(ii) large;

(iii) premium. [4]

The selling prices of the different categories of avocado at this supermarket are shown in the following table:

Category	Medium	Large	Premium
Selling price (\$) **per avocado**	1.10	1.29	1.96

The supermarket pays the farm \$200 for the avocados and assumes it will then sell them in exactly the same proportion as purchased from the farm.

(d) According to this model, find the minimum number of avocados that must be sold so that the net profit for the supermarket is at least \$438. [4]

Question 11

[Maximum mark: 7]

A factory manufactures lamps. It is known that the probability that a lamp is found to be defective is 0.05. A random sample of 30 lamps is tested.

(a) Find the probability that there is at least one defective lamp in the sample. [3]

(b) Given that there is at least one defective lamp in the sample, find the probability that there are at most two defective lamps. [4]

Question 12

[Maximum mark: 5]

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]

Question 13

[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 $\sigma\text{ g}$. 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]

Question 14

[Maximum mark: 6]

Let A and B be two independent events such that $\mathrm{P}(A \cap B') = 0.16$ and $\mathrm{P}(A' \cap B) = 0.36$.

(a) Given that $\mathrm{P}(A \cap B) = x$, find the value of x. [4]

(b) Find $\mathrm{P}(A' \mid B')$. [2]

Question 15

[Maximum mark: 6]

A discrete random variable, X, has the following probability distribution:

x	0	1	2	3
$\mathrm{P}(X=x)$	0.41	$k-0.28$	0.46	$0.29-2k^2$

(a) Show that $2k^2-k+0.12=0$. [1]

(b) Find the value of k, giving a reason for your answer. [3]

(c) Hence, find $\mathrm{E}(X)$. [2]

Question 16

[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]

Question 17

[Maximum mark: 18]

The time it takes Suzi to drive from home to work each morning is normally distributed with a mean of 35 minutes and a standard deviation of σ minutes.

On 25% of days, it takes Suzi longer than 40 minutes to drive to work.

(a) Find the value of σ. [4]

(b) On a randomly selected day, find the probability that Suzi's drive to work will take longer than 45 minutes. [2]

Suzi will be late to work if it takes her longer than 45 minutes to drive to work. The time it takes to drive to work each day is independent of any other day.

Suzi will work five days next week.

(c) Find the probability that she will be late to work at least one day next week. [3]

(d) Given that Suzi will be late to work at least one day next week, find the probability that she will be late less than three times. [5]

Suzi will work 22 days this month. She will receive a bonus if she is on time at least 20 of those days.

So far this month, she has worked 16 days and been on time 15 of those days.

(e) Find the probability that Suzi will receive a bonus. [4]

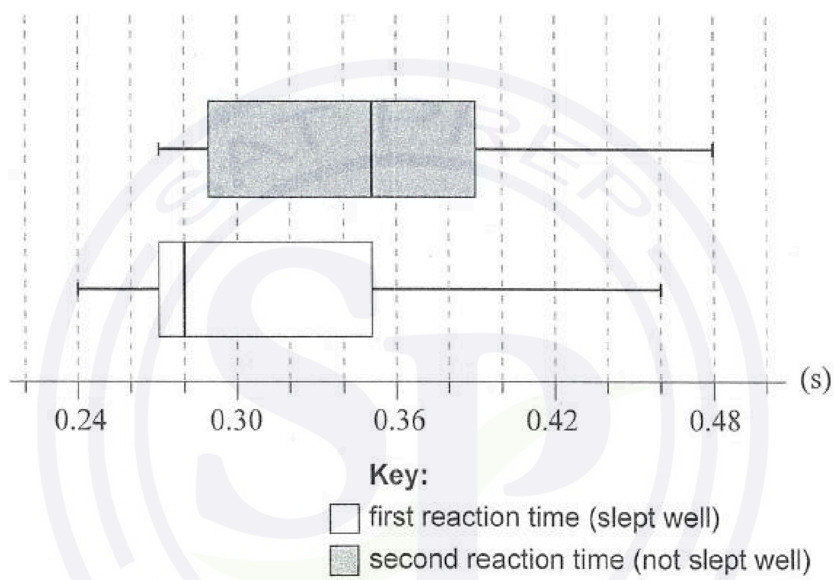
Question 18

[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]

Question 19

[Maximum mark: 6]

Events A and B are independent and $\mathrm{P}(A) = 3\mathrm{P}(B)$.

Given that $\mathrm{P}(A \cup B) = 0.68$, find $\mathrm{P}(B)$.

Question 20

[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 $\mathrm{P}(a \le T \le b) = 0.904$ and that $\mathrm{P}(T > b) = 2\mathrm{P}(T < a)$, 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]

Question 21

[Maximum mark: 5]

The following table shows the Mathematics test scores (x) and the Science test scores (y) for a group of eight students.

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

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) 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]

Question 22

[Maximum mark: 17]

In a large city, 160 people were surveyed. Of those, 60 were children (C) and the rest adults (A).

Each person in the survey was asked whether they preferred milk chocolate (M) or dark chocolate (D). It was found that 48 of the children preferred milk chocolate. This information is shown in the following table.

	M (milk chocolate)	**D (dark chocolate)**	**Total**
C (children)	48	p	60
A (adults)	x	y	q

(a) Find the value of

(i) p;

(ii) q. [2]

(b) Three people are chosen at random from those surveyed. Find the probability that all three are adults. [4]

(c) (i) Given that $\mathrm{P}(A \mid M) = \frac{1}{3}$, find the value of x.

(ii) A person is chosen at random from those surveyed. Write down the probability that they are an adult who prefers milk chocolate. [4]

(d) Determine if the events A and M are independent. Justify your answer. [3]

It can be assumed that the survey results are representative of the population of the city.

(e) Ten people in the city are chosen at random. Find the probability that at least five of them prefer dark chocolate. [4]

Question 23

[Maximum mark: 5]

A company manufactures metal tubes for bicycle frames. The diameters of the tubes, $D\,\mathrm{mm}$, are normally distributed with mean 32 and standard deviation σ. The interquartile range of the diameters is 0.28.

Find the value of σ.

Question 24

[Maximum mark: 7]

The total number of children, y, visiting a park depends on the highest temperature, T, in degrees Celsius ($°C$). A park official predicts the total number of children visiting his park on any given day using the model $y = -0.6T^2 + 23T + 110$, where $10 \le T \le 35$.

(a) Use this model to estimate the number of children in the park on a day when the highest temperature is $25\,°C$. [2]

An ice cream vendor investigates the relationship between the total number of children visiting the park and the number of ice creams sold, x. The following table shows the data collected on five different days.

Total number of children (y)	81	175	202	346	360
Ice creams sold (x)	15	27	23	35	46

(b) Find an appropriate regression equation that will allow the vendor to predict the number of ice creams sold on a day when there are y children in the park. [3]

(c) Hence, use your regression equation to predict the number of ice creams that the vendor sells on a day when the highest temperature is $25\,°C$. [2]

Question 25

[Maximum mark: 16]

A farmer is growing a field of rice 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 $\mathrm{P}(H < 82.4) = 0.213$ and $\mathrm{P}(H > 87.3) = 0.409$.

(a) Find the probability that the height of a randomly selected plant is between 82.4 cm and 87.3 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 87.3 cm is considered ready to harvest. Heights of plants are independent of each other.

(c) (i) Find the probability that exactly 32 plants are ready to harvest.

(ii) Given that fewer than 44 plants are ready to harvest, find the probability that exactly 32 plants are ready to harvest. [6]

In another field, the farmer is growing the same variety of rice, but is using a different fertilizer. The heights of these plants, F cm, are normally distributed with mean 92.8 and standard deviation d. The farmer finds the interquartile range to be 4.52 cm.

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

Question 26

[Maximum mark: 5]

The following table shows the probability distribution of a discrete random variable X, where $a, k \in \mathbb{R}^+$.

x	1	2	3	4
$\mathrm{P}(X = x)$	k	k^2	a	k^3

Given that $\mathrm{E}(X) = 2.6$, find the value of a.

Question 27

[Maximum mark: 15]

A bag contains n balls. It is known that ten of the balls are green, and the rest of the balls are red. Balls are drawn from the bag, one after the other, without replacement.

(a) Find, in terms of n, the probability that

(i) the first ball drawn is green;

(ii) the first two balls are green. [3]

For the following parts of this question, let $n = 25$.

(b) Show that the probability that the first two balls are red is 0.35. [2]

(c) Find the probability that the first three balls are all red. [2]

(d) Find the probability that at least one of the first three balls is green. [2]

A game is played where **four** balls are drawn, one after the other, from the bag of 25 balls, without replacement. A player earns points based on when the first green ball is drawn. At the end of each game, the four balls are put back in the bag.

A player earns zero points if no green ball is picked, or if the first green ball is picked on the first or second draw.

A player earns 10 points if the first green ball is picked on the third draw and earns 50 points if the first green ball is picked on the fourth draw.

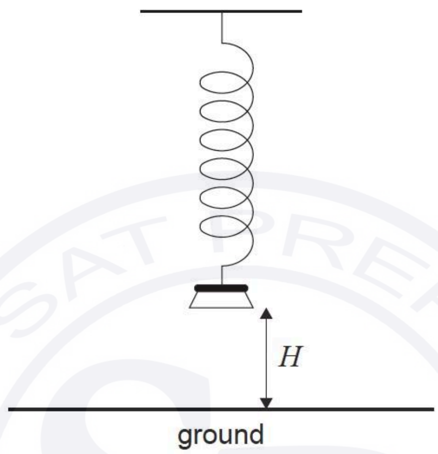
Millie plays this game k times. She finds her score by adding together her points from each game.

(e) Find the least value of k such that Millie's expected score is greater than 100. [6]

Question 28
[Maximum Mark: 13]

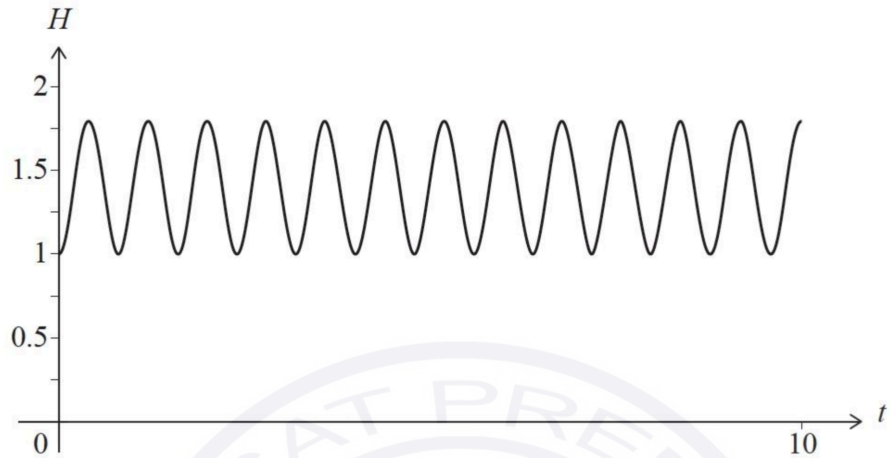
A weight suspended on a spring is pulled down and released, so that it moves up and down vertically.

The height, H metres, of the base of the weight above the ground can be modelled by the function $H(t) = a\cos(7.8t) + b$, for a, $b \in \mathbb{R}$ and $0 \le t \le 10$, where t is the time in seconds after the weight is released.



(a) Find the period of the function. [2]

The weight is released when its base is at a minimum height of 1 metre above the ground, and it reaches a maximum height of 1.8 metres above the ground. The graph of H is shown in the following diagram.



(b) Find the value of

(i) a;

(ii) b. [3]

(c) Find the number of times that the weight reaches its maximum height in the first five seconds of its motion. [2]

(d) Find the first time that the base of the weight reaches a height of 1.5 metres. [2]

A camera is set to take a picture of the weight at a random time during the first five seconds of its motion.

(e) Find the probability that the height of the base of the weight is greater than 1.5 metres at the time the picture is taken. [4]

Question 29

[Maximum mark: 8]

The weights, W grams, of bags of rice packaged in a factory can be modelled by a normal distribution with mean 204 grams and standard deviation 5 grams.

(a) A bag of rice is selected at random.

Find the probability that it weighs more than 210 grams. [2]

According to this model, 80% of the bags of rice weigh between w grams and 210 grams.

(b) Find the probability that a randomly selected bag of rice weighs less than w grams. [2]

(c) Find the value of w. [2]

(d) Ten bags of rice are selected at random.

Find the probability that exactly one of the bags weighs less than w grams. [2]

Question 30

[Maximum mark: 5]

A botanist is conducting an experiment which studies the growth of plants.

The heights of the plants are measured on seven different days.

The following table shows the number of days, d, that the experiment has been running and the average height, h cm, of the plants on each of those days.

Number of days (d)	2	5	13	24	33	37	42
Average height (h)	10	16	30	59	76	79	82

(a) The regression line of h on d for this data can be written in the form $h = ad + b$.

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 your regression line to estimate the average height of the plants when the experiment has been running for 20 days. [2]

Question 31

[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 $\mathrm{P}(H < 94.6) = 0.288$ and $\mathrm{P}(H > 98.1) = 0.434$.

(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]

Question 32

[Maximum mark: 5]

The following table shows the probability distribution of a discrete random variable X, where $a, k \in \mathbb{R}^+$.

x	1	2	3	4
$\mathrm{P}(X = x)$	k	k^2	a	k^3

Given that $\mathrm{E}(X) = 2.3$, find the value of a.

Question 33

[Maximum mark: 14]

A lake contains a type of fish called carp. The lengths, L cm, of the carp can be modelled by a normal distribution with mean 45.6 cm and standard deviation 4.2 cm.

According to this model, carp with a length between 41.4 cm and $k\text{ cm}$ lie within one standard deviation of the mean.

(a) Write down the value of k. [2]

(b) Find the probability that a randomly selected carp is greater than 48 cm in length. [2]

(c) It is known that 99% of carp in the lake have a length greater than $x\text{ cm}$. Find the value of x. [2]

(d) Consider a random sample of 100 carp from the lake.

(i) Find the expected number of carp with lengths between 40 cm and 56 cm.

(ii) Find the probability that in this sample, exactly 95 carp have a length between 40 cm and 56 cm. [5]

A large sample of carp from the lake is studied. The length of each fish is measured and recorded correct to the nearest 0.1 cm.

(e) Find the probability that a randomly selected carp has a length recorded as 45.6 cm. [3]

Question 34

[Maximum mark: 5]

Consider a random variable X such that $X \sim \mathrm{B}(n, 0.25)$.

Determine the least value of n such that $\mathrm{P}(X \geq 1) > 0.99$.

Question 35

[Maximum mark: 5]

Consider the following bivariate data set where $p, q \in \mathbb{Z}^+$.

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

The regression line of y on x has equation $y = 2.1875x + 0.6875$.

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]

Question 36

[Maximum mark: 6]

In a study, the mobile phone usage of a random sample of ten students was examined on a particular day.

The length of time, t hours, that the ten students used their phones for are listed below.

0.7 1.2 1.9 4.0 4.4 4.5 4.9 5.7 6.5 11.7

(a) For these data, find the

(i) median;

(ii) interquartile range. [3]

An outlier is a value that is less than $Q_1 - 1.5 \times \text{IQR}$ or greater than $Q_3 + 1.5 \times \text{IQR}$.

(b) Show that 11.7 is an outlier. [3]

Question 37

[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]

Question 38

[Maximum mark: 4]

The random variable X is normally distributed with mean 10 and standard deviation 2.

(a) Find the probability that X is more than 1.5 standard deviations above the mean. [2]

The probability that X is more than k standard deviations above the mean is 0.1, where $k \in \mathbb{R}$.

(b) Find the value of k. [2]

Question 39

[Maximum mark: 7]

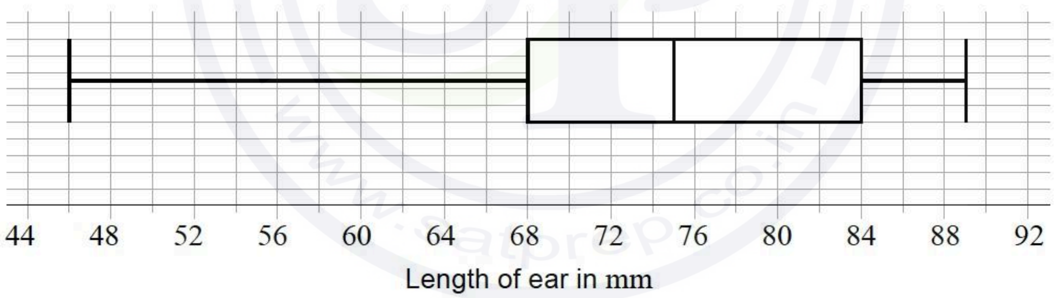
Janie claims that rabbits in Australia have longer ears than rabbits in Spain.

To test her claim, a randomly selected sample of rabbits was collected in each country.

The length of one ear of each rabbit was measured and the value recorded correct to the nearest millimetre (mm).

In the Australian sample, the median recorded value was $80\,\text{mm}$ and the interquartile range was $11\,\text{mm}$.

The recorded values for the Spanish sample are shown in the following box and whisker diagram.



(a) Complete the following table for the recorded values of the lengths of the rabbits' ears in each sample. [3]

	Australia	Spain
Median (mm)	80	
Interquartile range (mm)	11	

(b) Justifying your answers, compare the distributions of the lengths of rabbits' ears in Australia and Spain using

(i) the median;

(ii) the interquartile range. [4]

Question 40

[Maximum mark: 15]

The following table shows the population of Canada t years after the year 2000.

t (years after 2000)	0	5	10	15	20
p (population in millions)	30.7	32.2	34.0	35.7	37.9

A student uses linear regression to model the population of Canada using these data. The student model is $p = at + b$.

(a) (i) Write down the value of a and the value of b.

(ii) Interpret, in context, the value of a. [3]

The student uses this model to predict the population of Canada in the year 2030, where $t = 30$, and calculates a population of approximately 41.3 million people.

(b) Comment on the reliability of the student's prediction. [1]

A data scientist, Benoit, uses additional information to develop an exponential model for Canada's future population.

In this model, $B(t) = 33.5(1.005)^t$ represents the millions of people in Canada t years after the year 2000, where $25 \le t \le 100$.

(c) (i) Use Benoit's model to predict the population of Canada in the year 2100.

(ii) Interpret, in context, the value 1.005 in Benoit's model. [3]

Another data scientist, Cecilia, develops a third model for the Canadian population.

In this model, $C(t) = \dfrac{62}{1 + e^{-0.02t}}$ represents the millions of people in Canada t years after the year 2000, where $25 \le t \le 100$.

(d) Use Cecilia's model to predict the population of Canada in the year 2100. [1]

(e) Determine the year in which the difference between the predictions from Benoit's model and Cecilia's model is greatest. [3]

(f) Find the value of

(i) $B'(75)$;

(ii) $C'(75)$. [2]

(g) Compare and interpret, in context, the values of $B'(75)$ and $C'(75)$. [2]

Question 41

[Maximum mark: 14]

MyLife is a social media platform with 93.6 million users, all aged 13 years old and above.
The following frequency table shows the number of users by age group.

Age, a (years)	**Millions of users**
$13 \le a < 18$	5.5
$18 \le a < 25$	23.6
$25 \le a < 45$	43.5
$45 \le a < 65$	17.3
$65 \le a < 85$	3.7
Total	**93.6**

(a) Find the percentage of *MyLife* users aged 45 years or older. [1]

(b) A sample of 1000 *MyLife* users is chosen at random. Find the probability that fewer than 200 of them are aged 45 years or older. [3]

(c) Given that a *MyLife* user chosen at random is 45 years or older, find the probability that they are 65 years or older. [4]

(d) List the mid-interval value for each class interval. [1]

(e) Hence, for *MyLife* users, estimate

(i) the mean age;

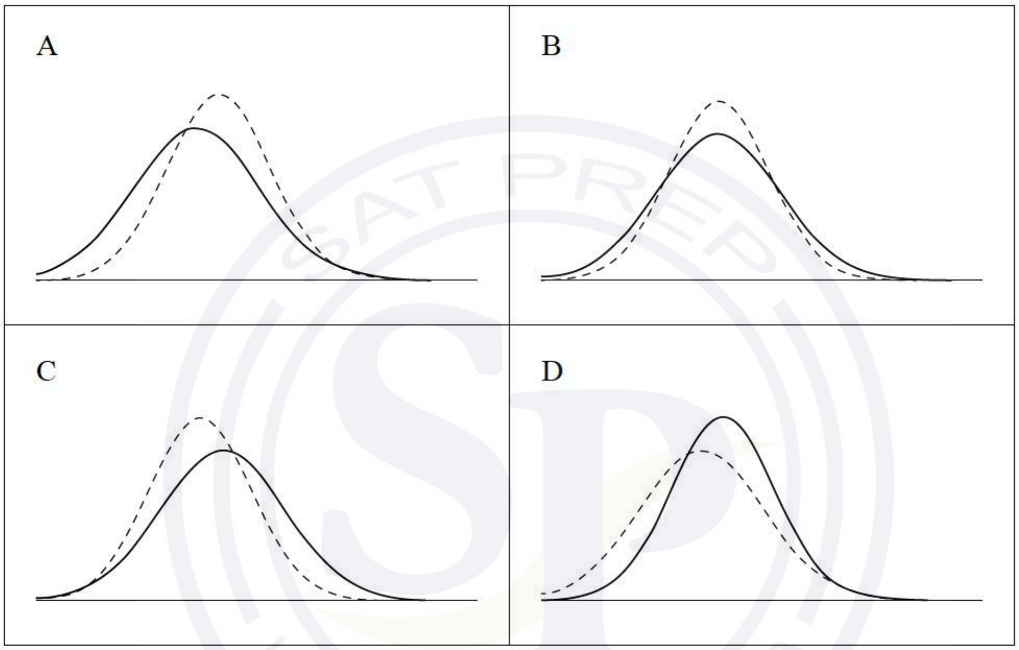
(ii) the variance of the ages. [3]

A different social media platform, *SmallTalk*, reports that its users have a mean age of 29.9 years and a variance of 137 years^2.

The following four diagrams represent age distributions.

(f) (i) Identify the diagram which best represents the age distributions for the users of *MyLife* and *SmallTalk*.

(ii) In your chosen diagram, identify which social media platform is represented by the solid line. [2]



Question 42

[Maximum mark: 5]

A discrete random variable, X, has the following probability distribution:

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

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

(b) Find $E(X)$. [3]

Question 43

[Maximum mark: 15]

The following table shows the population of Canada t years after the year 2000.

t (years after 2000)	0	5	10	15	20
p (population in millions)	30.6	32.3	34.1	35.6	38.0

A student uses linear regression to model the population of Canada using these data. The student model is $p = at + b$.

(a) (i) Write down the value of a and the value of b.

(ii) Interpret, in context, the value of a. [3]

The student uses this model to predict the population of Canada in the year 2030, where $t = 30$, and calculates a population of approximately 41.4 million people.

(b) Comment on the reliability of the student's prediction. [1]

A data scientist, Benoit, uses additional information to develop an exponential model for Canada's future population.

In this model, $B(t) = 30.6(1.007)^t$ represents the millions of people in Canada t years after the year 2000, where $25 \le t \le 100$.

(c) (i) Use Benoit's model to predict the population of Canada in the year 2100.

(ii) Interpret, in context, the value 1.007 in Benoit's model. [3]

Another data scientist, Cecilia, develops a third model for the Canadian population.

In this model, $C(t) = \frac{61}{1 + e^{-0.03t}}$ represents the millions of people in Canada t years after the year 2000, where $25 \le t \le 100$.

(d) Use Cecilia's model to predict the population of Canada in the year 2100. [1]

(e) Determine the year in which the difference between the predictions from Benoit's model and Cecilia's model is greatest. [3]

(f) Find the value of

(i) $B'(40)$;

(ii) $C'(40)$. [2]

(g) Compare and interpret, in context, the values of $B'(40)$ and $C'(40)$. [2]

Question 44

[Maximum mark: 14]

MyLife is a social media platform with 89.8 million users, all aged 12 years old and above. The following frequency table shows the number of users by age group.

Age, a (years)	**Millions of users**
$12 \le a < 18$	5.8
$18 \le a < 35$	42.7
$35 \le a < 55$	26.3
$55 \le a < 75$	12.9
$75 \le a < 95$	2.1
Total	**89.8**

(a) Find the percentage of *MyLife* users aged 55 years or older. [1]

(b) A sample of 1000 *MyLife* users is chosen at random. Find the probability that fewer than 150 of them are aged 55 years or older. [3]

(c) Given that a *MyLife* user chosen at random is 55 years or older, find the probability that they are 75 years or older. [4]

(d) List the mid-interval value for each class interval. [1]

(e) Hence, for *MyLife* users, estimate

(i) the mean age;

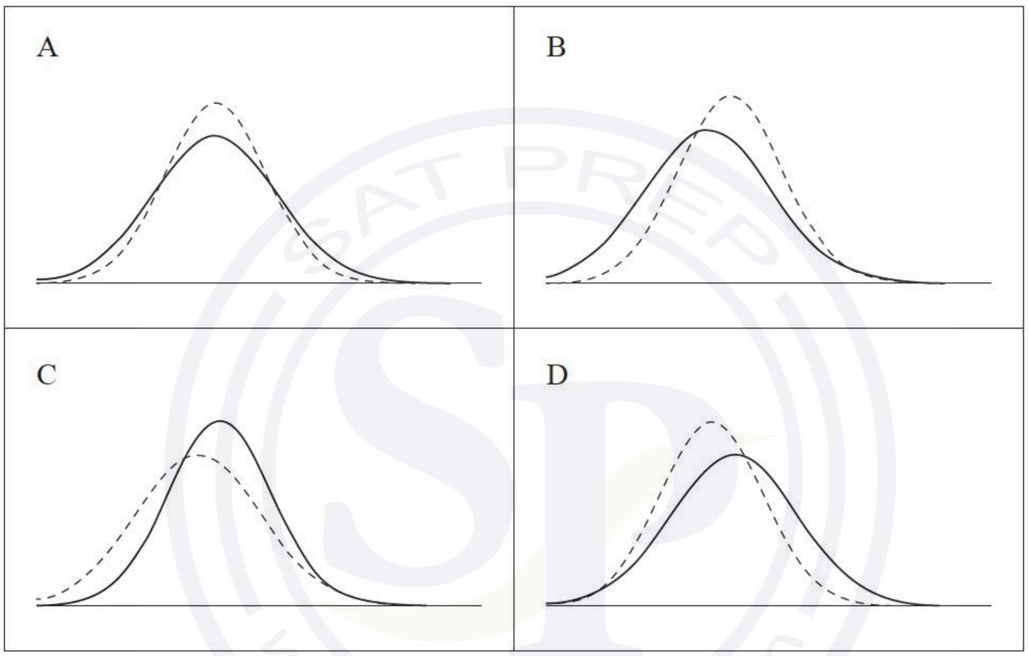
(ii) the variance of the ages. [3]

A different social media platform, *SmallTalk*, reports that its users have a mean age of 29.9 years and a variance of 137 years^2.

The following four diagrams represent age distributions.

(f) (i) Identify the diagram which best represents the age distributions for the users of *MyLife* and *SmallTalk*.

(ii) In your chosen diagram, identify which social media platform is represented by the dotted line. [2]



Question 45

[Maximum mark: 5]

A discrete random variable, X, has the following probability distribution:

$$\mathrm{P}(X = x) = \frac{kx}{15} \text{ for } x \in \{2, 4, 7, 10\}.$$

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

(b) Find $\mathrm{E}(X)$. [3]