Subject – Math AA(Higher 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: 6]

In a city, the number of passengers, X, who ride in a taxi has the following probability distribution.

x	1	2	3	4	5
$P(X=x)$	0.60	0.30	0.03	0.05	0.02

After the opening of a new highway that charges a toll, a taxi company introduces a charge for passengers who use the highway. The charge is $ 2.40 per taxi plus $ 1.20 per passenger. Let T represent the amount, in dollars, that is charged by the taxi company per ride.

(a) Find $E(T)$. [4]

(b) Given that $\text{Var}(X) = 0.8419$, find $\text{Var}(T)$. [2]

Question 3

[Maximum mark: 7]

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]

(c) Jennifer was absent for the first test but scored 29 marks on the second test. Use an appropriate regression equation to estimate Jennifer's mark on the first test. [2]

Question 4

[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 5

[Maximum mark: 7]

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

$$f(x)=\begin{cases}\dfrac{x}{\sqrt{\left(x^2+k\right)^3}} & 0\le x\le 4\\ 0 & \text{otherwise}\end{cases}$$

where $k\in\mathbb{R}^+$.

(a) Show that $\sqrt{16+k}-\sqrt{k}=\sqrt{k}\sqrt{16+k}$. [5]

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

Question 6

[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]

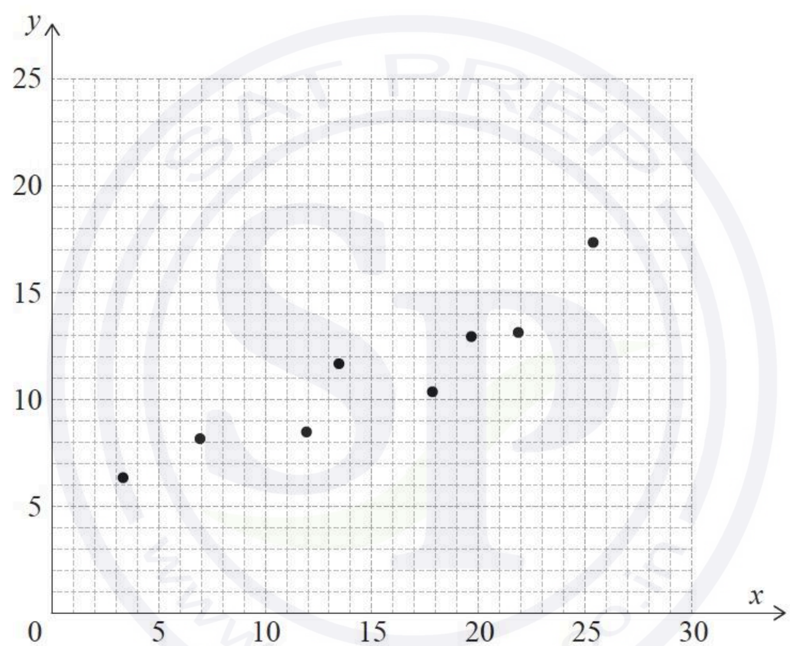
Question 7

[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 8

[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 9

[Maximum mark: 6]

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

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

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

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

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

Question 10

[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 11

[Maximum mark: 6]

A continuous random variable X has a probability density function given by

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

The median of this distribution is m.

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

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

Question 12

[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 13

[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 14

[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]

Question 15

[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 16

[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 17

[Maximum mark: 6]

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

x	0	1	2	3
$P(X=x)$	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 $E(X)$. [2]

Question 18

[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 19

[Maximum mark: 7]

Rachel and Sophia are competing in a javelin-throwing competition.

The distances, R metres, thrown by Rachel can be modelled by a normal distribution with mean 56.5 and standard deviation 3.

The distances, S metres, thrown by Sophia can be modelled by a normal distribution with mean 57.5 and standard deviation 1.8.

In the first round of competition, each competitor must have five throws. To qualify for the next round of competition, a competitor must record at least one throw of 60 metres or greater in the first round.

Find the probability that only one of Rachel or Sophia qualifies for the next round of competition.

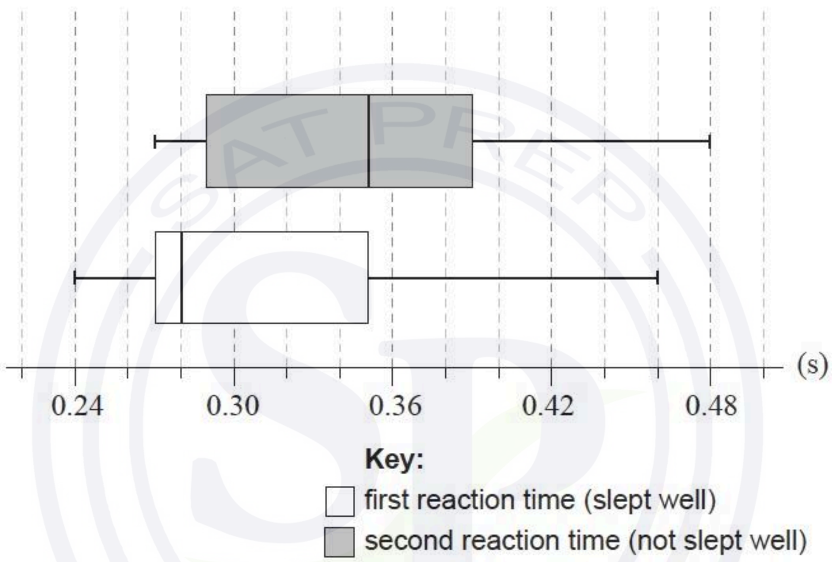
Question 20

[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 21

[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 22

[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 23

[Maximum mark: 8]

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

$$f(x) = \begin{cases} axe^{x}, & 0 \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

where $a, b \in \mathbb{R}^{+}$.

(a) Find an expression for a in terms of b. [5]

(b) In the case where $a = b = 1$, find the median of X. [3]

Question 24

[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 25

[Maximum mark: 19]

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

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

(a) Find the exact value of $\mathrm{E}(X)$. [5]

(b) Find $\mathrm{P}(X < 0.5)$. [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]

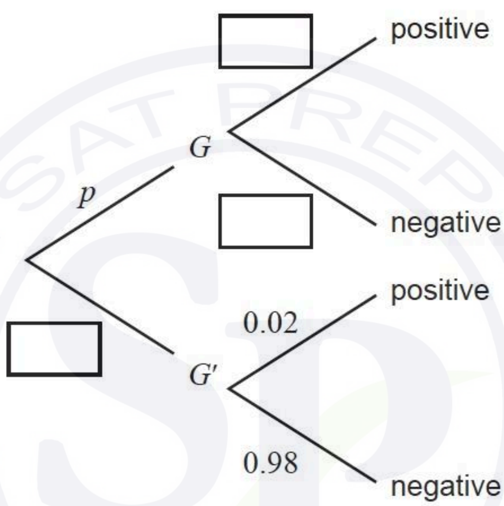
Question 26

[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]

Question 27

[Maximum mark: 5]

A company manufactures metal tubes for bicycle frames. The diameters of the tubes, $D\,\text{mm}$, 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 28

[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 $25\,°\text{C}$. [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 $25\,°\text{C}$. [2]

Question 29

[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 $\mathrm{P}(YY)$ represent the probability of drawing two yellow balls from the box without replacement.

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

(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 $\mathrm{P}(YY)=\frac{1}{3}$. [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 $\mathrm{P}(YYY)$ represent the probability of drawing three yellow balls from the box without replacement.

(d) Find an expression for $\mathrm{P}(YYY)$ 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]

Question 30

[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 31

[Maximum mark: 4]

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

The value of Pearson's product-moment correlation coefficient, r, for this data is 0.991, correct to three significant figures.

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

Question 32

[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]

Question 33

[Maximum mark: 5]

The random variable X is such that $X \sim B(25, p)$ and $\text{Var}(X) = 5.75$.

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

The random variable Y is such that $Y = 1 - 2X$.

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

Question 34

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

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

Question 35

[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

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

(a) Find the mode of X. [2]

(b) Find $\mathrm{P}(1\le X\le 2)$. [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]

Question 36

[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 $\mathrm{P}(X\ge 1)>0.99$.

Question 37

[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 38

[Maximum mark: 6]

A continuous random variable X has a probability density function f given by

$$f(x) = \begin{cases} kx & 0 \le x \le k \\ 2kx - x^2 & k < x \le 2k \\ 0, & \text{otherwise} \end{cases}$$

where $k > 0$.

(a) Show that k satisfies the equation $7k^3 = 6$. [2]

(b) Find the median of X. [4]

Question 39

[Maximum mark: 6]

In Happyland, the weather on any given day is independent of the weather on any other day. On any day in May, the probability of rain is 0.2. May has 31 days.

Find the probability that

(a) it rains on exactly 10 days in May; [2]

(b) it rains on at least 10 days in May; [2]

(c) the first day that it rains in May is on the 10th day. [2]

Question 40

[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 41

[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 42

[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 43

[Maximum mark: 8]

A continuous random variable X has probability density function

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

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

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

(c) Find the median of X. [3]

Question 44

[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]