Subject – Math AI(Higher Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2024
Paper -1
Questions

Question 1

[Maximum mark: 6]

Mr Burke teaches a mathematics class with 15 students. In this class there are 6 female students and 9 male students.

Each day Mr Burke randomly chooses one student to answer a homework question.

In the first month, Mr Burke will teach his class 20 times.

(a) Find the probability he will choose a female student 8 times. [2]

The Head of Year, Mrs Smith, decides to select a student at random from the year group to read the notices in assembly. There are 80 students in total in the year group. Mrs Smith calculates the probability of picking a male student 8 times in the first 20 assemblies is 0.153357 correct to 6 decimal places.

(b) Find the number of male students in the year group. [4]

Question 2

[Maximum mark: 6]

The number of fish that can be caught in one hour from a particular lake can be modelled by a Poisson distribution.

The owner of the lake, Emily, states in her advertising that the average number of fish caught in an hour is three.

Tom, a keen fisherman, is not convinced and thinks it is less than three. He decides to set up the following test. Tom will fish for one hour and if he catches fewer than two fish he will reject Emily's claim.

(a) State a suitable null and alternative hypotheses for Tom's test. [1]

(b) Find the probability of a Type I error. [2]

The average number of fish caught in an hour is actually 2.5.

(c) Find the probability of a Type II error. [3]

Question 3

[Maximum mark: 6]

In a coffee shop, the time it takes to serve a customer can be modelled by a normal distribution with a mean of 1.5 minutes and a standard deviation of 0.4 minutes.

Two customers enter the shop together. They are served one at a time.

Find the probability that the total time taken to serve both customers will be less than 4 minutes.

Clearly state any assumptions you have made.

Question 4

[Maximum mark: 5]

A manager wishes to check the mean weight of flour put into bags in his factory. He randomly samples 10 bags and finds the mean weight is 1.478 kg and the standard deviation of the sample is 0.0196 kg.

(a) Find s_{n-1} for this sample. [2]

(b) Find a 95 % confidence interval for the population mean, giving your answer to 4 significant figures. [2]

(c) The bags are labelled as being 1.5 kg weight. Comment on this claim with reference to your answer in part (b). [1]

Question 5

[Maximum mark: 6]

Jae Hee plays a game involving a biased six-sided die.
The faces of the die are labelled -3, -1, 0, 1, 2 and 5.
The score for the game, X, is the number which lands face up after the die is rolled.
The following table shows the probability distribution for X.

Score x	-3	-1	0	1	2	5
$P(X=x)$	$\frac{1}{18}$	p	$\frac{3}{18}$	$\frac{1}{18}$	$\frac{2}{18}$	$\frac{7}{18}$

(a) Find the exact value of p. [1]

Jae Hee plays the game once.

(b) Calculate the expected score. [2]

Jae Hee plays the game twice and adds the two scores together.

(c) Find the probability Jae Hee has a **total** score of -3. [3]

Question 6

[Maximum mark: 6]

At the end of a school day, the Headmaster conducted a survey asking students in how many classes they had used the internet.

The data is shown in the following table.

Number of classes in which the students used the internet	0	1	2	3	4	5	6
Number of students	20	24	30	k	10	3	1

(a) State whether the data is discrete or continuous. [1]

The mean number of classes in which a student used the internet is 2.

(b) Find the value of k. [4]

It was not possible to ask every person in the school, so the Headmaster arranged the student names in alphabetical order and then asked every 10th person on the list.

(c) Identify the sampling technique used in the survey. [1]

Question 7

[Maximum mark: 6]

The weights of apples from Tony's farm follow a normal distribution with mean 158 g and standard deviation 13 g. The apples are sold in bags that contain six apples.

(a) Find the mean weight of a bag of apples. [2]

(b) Find the standard deviation of the weights of these bags of apples. [2]

(c) Find the probability that a bag selected at random weighs more than 1 kg. [2]

Question 8

[Maximum mark: 7]

A factory, producing plastic gifts for a fast food restaurant's Jolly meals, claims that just 1% of the toys produced are faulty.

A restaurant manager wants to test this claim. A box of 200 toys is delivered to the restaurant. The manager checks all the toys in this box and four toys are found to be faulty.

(a) Identify the type of sampling used by the restaurant manager. [1]

The restaurant manager performs a one-tailed hypothesis test, at the 10% significance level, to determine whether the factory's claim is reasonable. It is known that faults in the toys occur independently.

(b) Write down the null and alternative hypotheses. [2]

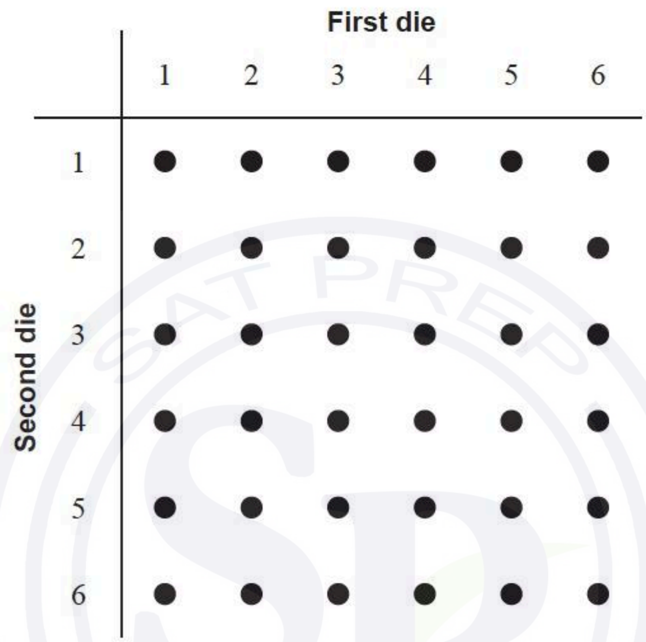
(c) Find the p-value for the test. [2]

(d) State the conclusion of the test. Give a reason for your answer. [2]

Question 9

[Maximum mark: 7]

A game is played where two unbiased dice are rolled and the score in the game is the greater of the two numbers shown. If the two numbers are the same, then the score in the game is the number shown on one of the dice. A diagram showing the possible outcomes is given below.



Let T be the random variable "the score in a game".

(a) Complete the table to show the probability distribution of T. [2]

t	1	2	3	4	5	6
$P(T = t)$						

(b) Find the probability that

(i) a player scores at least 3 in a game.

(ii) a player scores 6, given that they scored at least 3. [3]

(c) Find the expected score of a game. [2]

Question 10

[Maximum mark: 4]

George goes fishing. From experience he knows that the mean number of fish he catches per hour is 1.1. It is assumed that the number of fish he catches can be modelled by a Poisson distribution.

On a day in which George spends 8 hours fishing, find the probability that he will catch more than 9 fish.

Question 11

[Maximum mark: 7]

The number of coffees sold per hour at an independent coffee shop is modelled by a Poisson distribution with a mean of 22 coffees per hour.

Sheila, the shop's owner wants to increase the number of coffees sold in the shop. She decides to offer a discount to customers who buy more than one coffee.

To test how successful this strategy is, Sheila records the number of coffees sold over a single 5-hour period. Sheila decides to use a 5% level of significance in her test.

(a) State the null and alternative hypotheses for the test. [1]

(b) Find the probability that Sheila will make a type I error in her test conclusion. [4]

Sheila finds 126 coffees were sold during the 5-hour period.

(c) State Sheila's conclusion to the test. Justify your answer. [2]

Question 12

[Maximum mark: 6]

A manufacturer of chocolates produces them in individual packets, claiming to have an average of 85 chocolates per packet.

Talha bought 30 of these packets in order to check the manufacturer's claim.

Given that the number of individual chocolates is x, Talha found that, from his 30 packets, $\sum x = 2506$ and $\sum x^2 = 209\,738$.

(a) Find an unbiased estimate for the mean number (μ) of chocolates per packet. [1]

(b) Use the formula $s_{n-1}^2 = \dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n-1}$ to determine an unbiased estimate for the variance of the number of chocolates per packet. [2]

(c) Find a 95% confidence interval for μ. You may assume that all conditions for a confidence interval have been met. [2]

(d) Suggest, with justification, a valid conclusion that Talha could make. [1]

Question 13

. [Maximum mark: 8]

A newspaper vendor in Singapore is trying to predict how many copies of *The Straits Times* they will sell. The vendor forms a model to predict the number of copies sold each weekday. According to this model, they expect the same number of copies will be sold each day.

To test the model, they record the number of copies sold each weekday during a particular week. This data is shown in the table.

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Number of copies sold	74	97	91	86	112

A goodness of fit test at the 5% significance level is used on this data to determine whether the vendor's model is suitable. The critical value for the test is 9.49.

(a) Find an estimate for how many copies the vendor expects to sell each day. [1]

(b) (i) State the null and alternative hypotheses for this test.

(ii) Write down the degrees of freedom for this test.

(iii) Write down the conclusion to the test. Give a reason for your answer. [7]

Question 14

[Maximum mark: 6]

At Springfield University, the weights, in kg, of 10 chinchilla rabbits and 10 sable rabbits were recorded. The aim was to find out whether chinchilla rabbits are generally heavier than sable rabbits. The results obtained are summarized in the following table.

Weight of chinchilla rabbits, kg	4.9	4.2	4.1	4.4	4.3	4.6	4.0	4.7	4.5	4.4
Weight of sable rabbits, kg	4.2	4.1	4.1	4.2	4.5	4.4	4.5	3.9	4.2	4.0

A t-test is to be performed at the 5% significance level.

(a) Write down the null and alternative hypotheses. [2]

(b) Find the p-value for this test. [2]

(c) Write down the conclusion to the test. Give a reason for your answer. [2]

Question 15

[Maximum mark: 7]

On Paul's farm, potatoes are packed in sacks labelled $50\,\text{kg}$. The weights of the sacks of potatoes can be modelled by a normal distribution with mean weight $49.8\,\text{kg}$ and standard deviation $0.9\,\text{kg}$.

(a) Find the probability that a sack is under its labelled weight. [2]

(b) Find the lower quartile of the weights of the sacks of potatoes. [2]

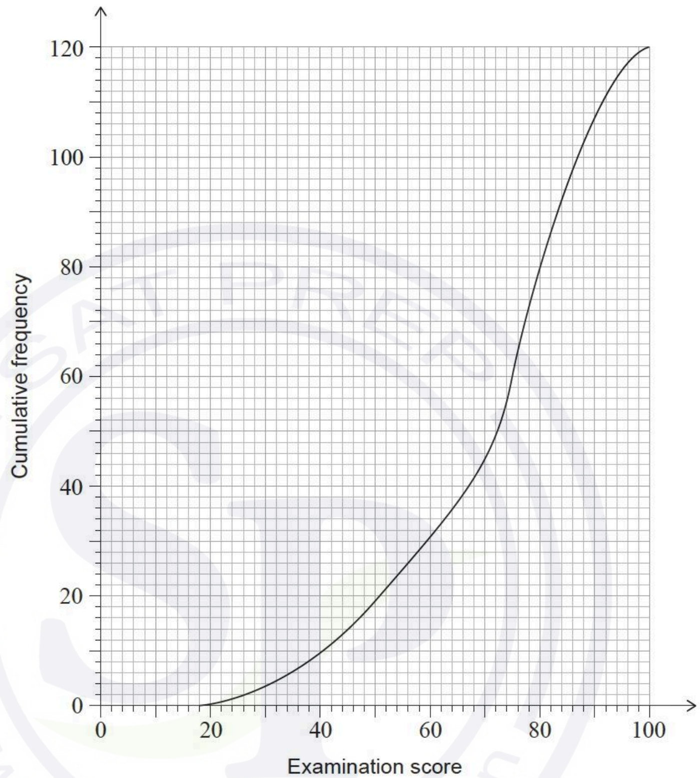
The sacks of potatoes are transported in crates. There are 10 sacks in each crate and the weights of the sacks of potatoes are independent of each other.

(c) Find the probability that the total weight of the sacks of potatoes in a crate exceeds $500\,\text{kg}$. [3]

Question 16

[Maximum mark: 8]

A group of 120 students sat a history exam. The cumulative frequency graph shows the scores obtained by the students.



(a) Find the median of the scores obtained. [1]

The students were awarded a grade from 1 to 5, depending on the score obtained in the exam. The number of students receiving each grade is shown in the following table.

Grade	1	2	3	4	5
Number of students	6	13	26	a	b

(b) Find an expression for a in terms of b. [2]

(c) The mean grade for these students is 3.65.

(i) Find the number of students who obtained a grade 5.

(ii) Find the minimum score needed to obtain a grade 5. [5]

Question 17

[Maximum mark: 7]

The number of cars arriving at a junction in a particular town in any given minute between 9:00 am and 10:00 am is historically known to follow a Poisson distribution with a mean of 5.4 cars per minute.

A new road is built near the town. It is claimed that the new road has decreased the number of cars arriving at the junction.

To test the claim, the number of cars, X, arriving at the junction between 9:00 am and 10:00 am on a particular day will be recorded. The test will have the following hypotheses:

H_0: the mean number of cars arriving at the junction has not changed,
H_1: the mean number of cars arriving at the junction has decreased.

The alternative hypothesis will be accepted if $X \leq 300$.

(a) Assuming the null hypothesis to be true, state the distribution of X. [1]

(b) Find the probability of a Type I error. [2]

(c) Find the probability of a Type II error, if the number of cars now follows a Poisson distribution with a mean of 4.5 cars per minute. [4]

Question 18

[Maximum mark: 8]

The strength of earthquakes is measured on the Richter magnitude scale, with values typically between 0 and 8 where 8 is the most severe.

The Gutenberg–Richter equation gives the average number of earthquakes per year, N, which have a magnitude of at least M. For a particular region the equation is

$$\log_{10} N = a - M, \text{ for some } a \in \mathbb{R}.$$

This region has an average of 100 earthquakes per year with a magnitude of at least 3.

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

The equation for this region can also be written as $N = \frac{b}{10^M}$.

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

Within this region the most severe earthquake recorded had a magnitude of 7.2.

(c) Find the average number of earthquakes in a year with a magnitude of at least 7.2. [1]

The number of earthquakes in a given year with a magnitude of at least 7.2 can be modelled by a Poisson distribution, with mean N. The number of earthquakes in one year is independent of the number of earthquakes in any other year.

Let Y be the number of years between the earthquake of magnitude 7.2 and the next earthquake of at least this magnitude.

(d) Find $\mathrm{P}(Y > 100)$. [3]

Question 19

[Maximum mark: 6]

A company produces bags of sugar with a labelled weight of $1\,\mathrm{kg}$. The weights of the bags are normally distributed with a mean of $1\,\mathrm{kg}$ and a standard deviation of $100\,\mathrm{g}$. In an inspection, if the weight of a randomly chosen bag is less than $950\,\mathrm{g}$ then the company fails the inspection.

(a) Find the probability that the company fails the inspection. [2]

A statistician in the company suggests it would be fairer if the company passes the inspection when the mean weight of five randomly chosen bags is greater than $950\,\mathrm{g}$.

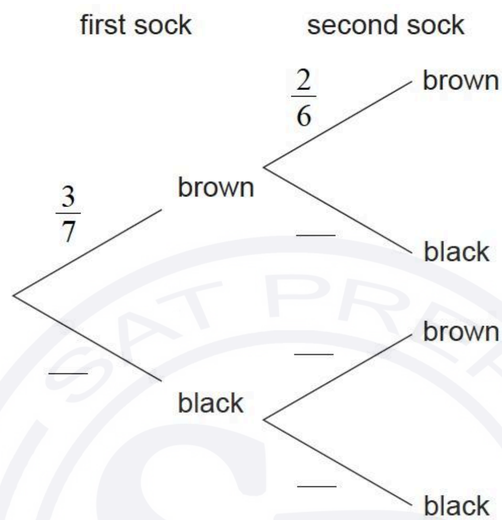
(b) Find the probability of passing the inspection if the statistician's suggestion is followed. [4]

Question 20

[Maximum mark: 7]

Karl has three brown socks and four black socks in his drawer. He takes two socks at random from the drawer.

(a) Complete the tree diagram. [2]



(b) Find the probability that Karl takes two socks of the same colour. [2]

(c) Given that Karl has two socks of the same colour find the probability that he has two brown socks. [3]

Question21

[Maximum mark: 7]

Leo is investigating whether a six-sided die is fair. He rolls the die 60 times and records the observed frequencies in the following table:

Number on die	1	2	3	4	5	6
Observed frequency	8	7	6	15	12	12

Leo carries out a χ^2 goodness of fit test at a 5% significance level.

(a) Write down the null and alternative hypotheses. [1]

(b) Write down the degrees of freedom. [1]

(c) Write down the expected frequency of rolling a 1. [1]

(d) Find the p-value for the test. [2]

(e) State the conclusion of the test. Give a reason for your answer. [2]

Question 22

[Maximum mark: 6]

The sex of cuttlefish is difficult to determine visually, so it is often found by weighing the cuttlefish.

The weights of adult male cuttlefish are known to be normally distributed with mean $10\,\text{kg}$ and standard deviation $0.5\,\text{kg}$.

The weights of adult female cuttlefish are known to be normally distributed with mean $12\,\text{kg}$ and standard deviation $1\,\text{kg}$.

A zoologist uses the null hypothesis that in the absence of information a cuttlefish is male.

If the weight is found to be above $11.5\,\text{kg}$ the cuttlefish is classified as female.

(a) Find the probability of making a Type I error when weighing a male cuttlefish. [2]

(b) Find the probability of making a Type II error when weighing a female cuttlefish. [2]

90% of adult cuttlefish are male.

(c) Find the probability of making an error using the zoologist's method. [2]

Question 23

[Maximum mark: 8]

A psychologist records the number of digits (d) of π that a sample of IB Mathematics higher level candidates could recall.

d	2	3	4	5	6	7
Frequency	2	6	24	21	11	3

(a) Find an unbiased estimate of the population mean of d. [1]

(b) Find an unbiased estimate of the population variance of d. [2]

The psychologist has read that in the general population people can remember an average of 4.4 digits of π. The psychologist wants to perform a statistical test to see if IB Mathematics higher level candidates can remember more digits than the general population.

(c) H_0: $\mu = 4.4$ is the null hypothesis for this test.

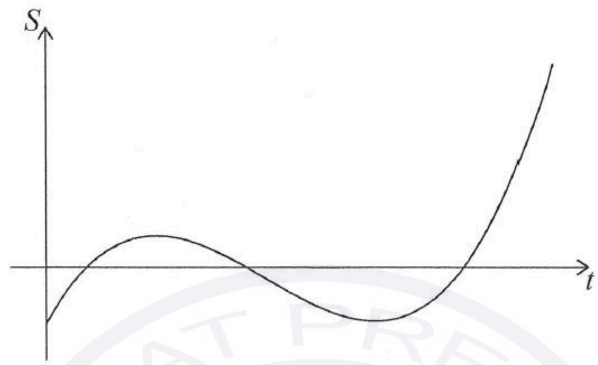
(i) State the alternative hypothesis.

(ii) Given that all assumptions for this test are satisfied, carry out an appropriate hypothesis test. State and justify your conclusion. Use a 5% significance level. [5]

Question 24

[Maximum mark: 8]

The graph below shows the average savings, S thousand dollars, of a group of university graduates as a function of t, the number of years after graduating from university.



(a) Write down one feature of this graph which suggests a cubic function might be appropriate to model this scenario. [1]

The equation of the model can be expressed in the form $S = at^3 + bt^2 + ct + d$, where a, b, c and d are real constants.

The graph of the model must pass through the following four points.

t	0	1	2	3
S	-5	3	-1	-5

(b) (i) Write down the value of d.

(ii) Write down three simultaneous equations for a, b and c.

(iii) Hence, or otherwise, find the values of a, b and c. [4]

A negative value of S indicates that a graduate is expected to be in debt.

(c) Use the model to determine the total length of time, in years, for which a graduate is expected to be in debt after graduating from university. [3]

Question 25

[Maximum mark: 7]

A polygraph test is used to determine whether people are telling the truth or not, but it is not completely accurate. When a person tells the truth, they have a 20% chance of failing the test. Each test outcome is independent of any previous test outcome.

10 people take a polygraph test and all 10 tell the truth.

(a) Calculate the expected number of people who will pass this polygraph test. [2]

(b) Calculate the probability that exactly 4 people will fail this polygraph test. [2]

(c) Determine the probability that fewer than 7 people will pass this polygraph test. [3]

Question 26

[Maximum mark: 6]

A group of 130 applicants applied for admission into either the Arts programme or the Sciences programme at a university. The outcomes of their applications are shown in the following table.

	Accepted	**Rejected**
Arts programme	17	24
Sciences programme	25	64

(a) Find the probability that a randomly chosen applicant from this group was accepted by the university. [1]

An applicant is chosen at random from this group. It is found that they were accepted into the programme of their choice.

(b) Find the probability that the applicant applied for the Arts programme. [2]

Two different applicants are chosen at random from the original group.

(c) Find the probability that both applicants applied to the Arts programme. [3]

Question 27

[Maximum mark: 8]

The time of sunrise, R hours after midnight, in Taipei can be modelled by

$$R = 1.08\cos(0.0165t + 0.413) + 4.94,$$

where t is the day of the year 2021 (for example, $t = 2$ represents 2 January 2021).

The time of sunset, S hours after midnight, in Taipei can be modelled by

$$S = 1.15\cos(0.0165t - 2.97) + 18.9.$$

The number of daylight hours, D, in Taipei during 2021 can be modelled by

$$D = a\cos(0.0165t + b) + c.$$

(a) Find the value of a, of b and of c. [6]

(b) Hence, or otherwise, find the largest number of daylight hours in Taipei during 2021 and the day of the year on which this occurs. [2]

Question 28

[Maximum mark: 8]

The principal of a school is concerned that only 30% of her students are choosing healthy options from the school canteen. She organizes a campaign to promote healthy eating and decides to test if the campaign has increased the number of students choosing healthy options. She assumes that a student's choice is independent of other students' choices.

(a) Write down suitable hypotheses for this test. [2]

The principal decides to take a random sample of 80 students. She will reject the null hypothesis if at least 31 students choose a healthy option.

(b) Find the probability that she makes a Type I error. [3]

In fact, the campaign led to 40% of her students choosing a healthy option.

(c) Find the probability that she makes a Type II error. [3]

Question 29

[Maximum mark: 7]

Taizo plays a game where he throws one ball at two bottles that are sitting on a table.
The probability of knocking over bottles, in any given game, is shown in the following table.

Number of bottles knocked over	0	1	2
Probability	0.5	0.4	0.1

(a) Taizo plays two games that are independent of each other. Find the probability that Taizo knocks over a **total** of two bottles. [4]

In any given game, Taizo will win k points if he knocks over two bottles, win 4 points if he knocks over one bottle and lose 8 points if no bottles are knocked over.

(b) Find the value of k such that the game is fair. [3]

Question 30

[Maximum mark: 5]

Sergio is interested in whether an adult's favourite breakfast berry depends on their income level. He obtains the following data for 341 adults and decides to carry out a χ^2 test for independence, at the 10% significance level.

		Income level		
		Low	**Medium**	**High**
Favourite berry	**Strawberry**	21	39	30
	Blueberry	39	67	42
	Other berry	32	45	26

(a) Write down the null hypothesis. [1]

(b) Find the value of the χ^2 statistic. [2]

The critical value of this χ^2 test is 7.78.

(c) Write down Sergio's conclusion to the test in context. Justify your answer. [2]

Question 31

[Maximum mark: 6]

The relationship between the intensity, I, of a light source and the distance, d, from the light source can be modelled by $I = \frac{k}{d^2}$.

Pablo measures the intensity of a light source at different distances. The data collected is shown in the table.

d (m)	1	2	5
I (lm)	42	11	1.5

Pablo finds the sum of square residuals in the form $1.0641k^2 - 89.62k + c$.

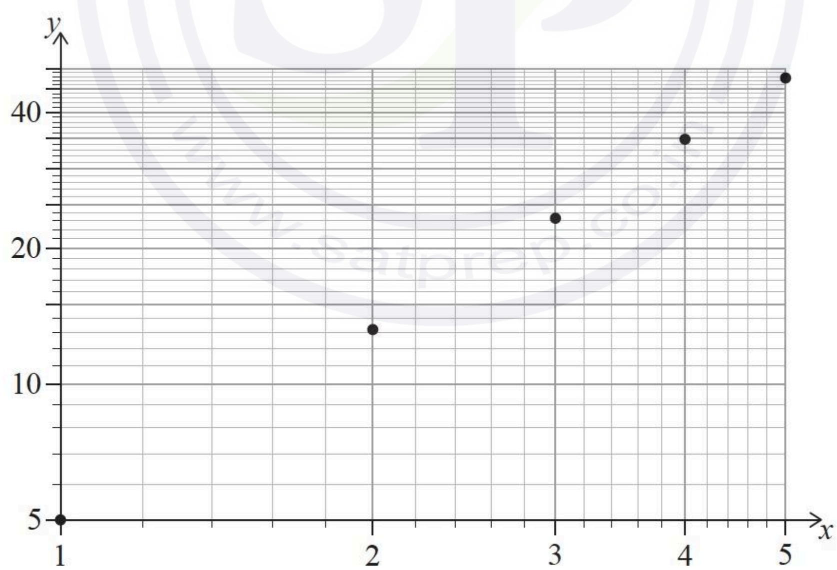
(a) Find the exact value of c. [4]

(b) Hence find the least squares regression curve of the form $I = \frac{k}{d^2}$. [2]

Question 32

[Maximum mark: 6]

Petra examines two quantities, x and y, and plots data points on a log-log graph.



She observes that on this graph the data points follow a perfect straight line. Given that the line passes through the points $(2, 13.1951)$ and $(4, 34.822)$, find the equation of the relationship connecting x and y. Your final answer should not include logarithms.

Question 33

[Maximum mark: 6]

A shop sells oranges and lemons. The weights of the oranges are assumed to be normally distributed with mean 205 grams and standard deviation 5 grams. The weights of the lemons are assumed to be normally distributed with mean 105 grams and standard deviation 3 grams.

Nelia selects 1 orange and 2 lemons at random and independent of each other. Calculate the probability that the weight of her orange exceeds the combined weight of her lemons.

Question 34

[Maximum mark: 5]

On a specific day, the speed of cars as they pass a speed camera can be modelled by a normal distribution with a mean of $67.3\,\text{km h}^{-1}$.

A speed of $75.7\,\text{km h}^{-1}$ is two standard deviations from the mean.

(a) Find the standard deviation for the speed of the cars. [2]

It is found that 82% of cars on this road travel at speeds between $p\,\text{km h}^{-1}$ and $q\,\text{km h}^{-1}$, where $p < q$. This interval includes cars travelling at a speed of $74\,\text{km h}^{-1}$.

(b) Show that the region of the normal distribution between p and q is **not** symmetrical about the mean. [3]

Question 35

[Maximum mark: 7]

The random variables (X, Y) follow a bivariate normal distribution with product-moment correlation coefficient ρ. The values of six random observations of (X, Y) are shown in the table.

x	6.3	4.1	5.6	9.2	7.8	8.2
y	9.2	4.9	8.9	10.3	8.9	9.8

(a) State null and alternative hypotheses which could be used to test whether there is a linear correlation between X and Y. [2]

(b) Determine the value of

(i) the product-moment correlation coefficient, r, of the sample.

(ii) the corresponding p-value. [3]

(c) State whether your result from part (b)(ii) indicates there is sufficient evidence to claim that, at the 5% significance level, X and Y are not linearly correlated.

Give a reason for your answer. [2]

Question 36

[Maximum mark: 6]

Carys believes that, on a memory retention test, the mean score of bilingual people (μ_b) will be higher than the mean score of monolingual people (μ_m). Carys gave a memory retention test to a random sample of students in her class. The results are shown in the two tables.

	Scores									
Bilingual	100	94	100	90	100	94	98	98	98	98

	Scores							
Monolingual	97	92	88	98	88	94	100	100

Carys performs a one-tailed t-test at a 5% level of significance. It is assumed that the scores are normally distributed and the samples have equal variances.

(a) State the null and alternative hypotheses. [2]

(b) Calculate the p-value for this test. [2]

(c) State the conclusion of the test in the context of the question. Justify your answer. [2]

Question 37

[Maximum mark: 6]

A random sample of eight packets of Apollo coffee granules are selected from a supermarket shelf.

The weights of the coffee granules present in each packet are as follows:

$$222\,\text{g} \quad 226\,\text{g} \quad 221\,\text{g} \quad 228\,\text{g} \quad 227\,\text{g} \quad 225\,\text{g} \quad 222\,\text{g} \quad 223\,\text{g}$$

(a) (i) Find an unbiased estimate for the mean weight of coffee granules in a packet of Apollo coffee.

(ii) Calculate a 95% confidence interval for the population mean. Give your answer to four significant figures. [3]

(b) State one assumption you have made in order for your interval to be valid. [1]

(c) The label of each packet has a description which includes the phrase: “contains $226\,\text{g}$ of coffee granules”.

Using your answer to part (a)(ii), briefly comment on the claim on the label. [2]

Question 38

[Maximum mark: 6]

A chocolate company plans to produce chocolate bars with special flavours. They survey 246 people to determine if there is any particular preference for one of the flavours.

The table below shows the information collected.

Hot chilli	Almond crunch	Spiced Chai	Ginger’n’lime
75	59	46	66

A χ^2 goodness of fit test at the 5% significance level is carried out on the data.

The critical value for the test is 7.82.

(a) State the null and alternative hypotheses for this test. [2]

(b) Perform the test and give your conclusion in context. [4]

Question 39

[Maximum mark: 9]

At a running club, Sung-Jin conducts a test to determine if there is any association between an athlete's age and their best time taken to run $100\,\text{m}$. Eight athletes are chosen at random, and their details are shown below.

Athlete	A	B	C	D	E	F	G	H
Age (years)	13	17	22	18	19	25	11	36
Time (seconds)	13.4	14.6	13.4	12.9	12.0	11.8	17.0	13.1

Sung-Jin decides to calculate the Spearman's rank correlation coefficient for his set of data.

(a) Complete the table of ranks. [2]

Athlete	A	B	C	D	E	F	G	H
Age rank			3					
Time rank							1	

(b) Calculate the Spearman's rank correlation coefficient, r_s. [2]

(c) Interpret this value of r_s in the context of the question. [1]

(d) Suggest a mathematical reason why Sung-Jin may have decided not to use Pearson's product-moment correlation coefficient with his data from the original table. [1]

(e) (i) Find the coefficient of determination for the data from the original table.

(ii) Interpret this value in the context of the question. [3]

Question 40

[Maximum mark: 6]

Akar starts a new job in Australia and needs to travel daily from Wollongong to Sydney and back. He travels to work for 28 consecutive days and therefore makes 56 single journeys. Akar makes all journeys by bus.

The probability that he is successful in getting a seat on the bus for any single journey is 0.86.

(a) Determine the expected number of these 56 journeys for which Akar gets a seat on the bus. [1]

(b) Find the probability that Akar gets a seat on at least 50 journeys during these 28 days. [3]

The probability that Akar gets a seat on at most n journeys is at least 0.25.

(c) Find the smallest possible value of n. [2]

Question 41

[Maximum mark: 6]

The lengths of the seeds from a particular mango tree are approximated by a normal distribution with a mean of $4\,\text{cm}$ and a standard deviation of $0.25\,\text{cm}$.

A seed from this mango tree is chosen at random.

(a) Calculate the probability that the length of the seed is less than $3.7\,\text{cm}$. [2]

It is known that $30\,\%$ of the seeds have a length greater than $k\,\text{cm}$.

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

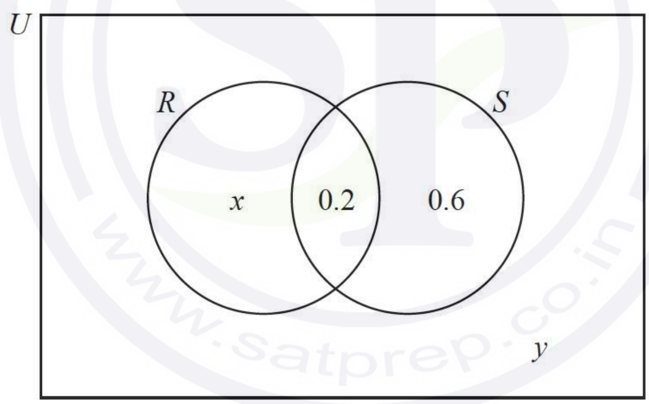
For a seed of length $d\,\text{cm}$, chosen at random, $\text{P}(4 - m < d < 4 + m) = 0.6$.

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

Question 42

[Maximum mark: 7]

The following Venn diagram shows two independent events, R and S. The values in the diagram represent probabilities.



(a) Find the value of x. [3]

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

(c) Find $\text{P}(R' \mid S')$. [2]

Question 43

[Maximum mark: 4]

A company that owns many restaurants wants to determine if there are differences in the quality of the food cooked for three different meals: breakfast, lunch and dinner.

Their quality assurance team randomly selects 500 items of food to inspect. The quality of this food is classified as perfect, satisfactory, or poor. The data is summarized in the following table.

		Quality			
		Perfect	**Satisfactory**	**Poor**	**Total**
Meal	**Breakfast**	101	124	7	**232**
	Lunch	68	81	5	**154**
	Dinner	35	69	10	**114**
	Total	**204**	**274**	**22**	**500**

A χ^2 test at the 5% significance level is carried out to determine if there is significant evidence of a difference in the quality of the food cooked for the three meals.

The critical value for this test is 9.488.

The hypotheses for this test are:

H_0: The quality of the food and the type of meal are independent.

H_1: The quality of the food and the type of meal are not independent.

(a) Find the χ^2 statistic. [2]

(b) State, with justification, the conclusion for this test. [2]

Question 44

[Maximum mark: 6]

When Jef plays basketball, the number of shots he takes during any 6 minutes of play can be modelled by a Poisson distribution with mean 2.5.

(a) Find the probability that Jef takes less than 7 shots during any 12 minutes of play. [2]

It can be assumed that the outcomes of the shots are independent of each other, and the probability of success of a shot is constant. The probability that Jef is successful with a shot is 0.4.

It can be assumed that the probability of Jef's success with a shot is independent of the number of shots that he takes.

(b) Find the probability that during any 6 minutes of play Jef takes fewer than 4 shots and is successful at least once. [4]

Question 45

[Maximum mark: 7]

The decay of a chemical isotope over five years is recorded in **Table 1**. The mass of the chemical M is measured to the nearest gram at the beginning of each year t of the experiment.

Table 1

Time t (years)	1	2	3	4	5
Mass M (grams)	1000	660	517	435	381

It is believed that the decay of the isotope can be modelled by an equation of the form $M = a \times t^b$.

(a) Use power regression on your graphic display calculator to find the value of a and the value of b. [2]

The values of t and M can be transformed such that $x = \ln t$ and $y = \ln M$. **Table 2** shows data for x and y to three decimal places.

Table 2

x	0	0.693	1.099	1.386	1.609
y	6.908	6.492	6.248	6.075	5.943

(b) Find the linear regression equation of y on x, in the form $y = cx + d$. Give the values of c and d to three decimal places. [2]

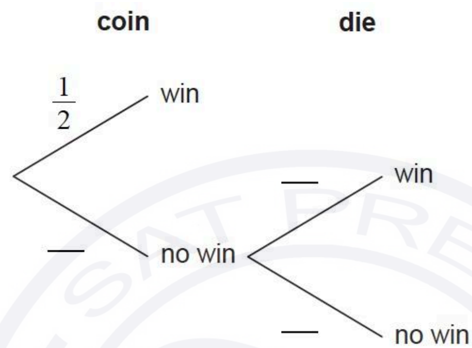
(c) Hence, show that this linear regression is equivalent to the power regression found in part (a). [3]

Question 46

[Maximum mark: 7]

Michèle is playing a game. In the game, she must first flip a fair coin which will result in the coin landing on heads or tails. If the coin lands on heads, then she wins a prize. If it lands on tails, then she has another chance but this time she must roll a fair six-sided dice and get a five or six in order to win a prize.

(a) Complete the tree diagram by writing in the three missing probabilities. [2]



(b) Find the probability that Michèle does **not** win a prize. [2]

(c) Given that Michèle won a prize, find the probability that the coin landed on heads. [3]

Question 47

[Maximum mark: 6]

A cup of hot water is placed in a room and is left to cool for half an hour. Its temperature, measured in $^\circ C$, is recorded every 5 minutes. The results are shown in the table.

Time (mins)	5	10	15	20	25	30
Temperature (°C)	59	52	46	40	35	29

Akira uses the power function $T(t) = at^b + 25$ to model the temperature, T, of the water t minutes after it was placed in the room.

(a) State what the value of 25 represents in this context. [1]

(b) Use your graphic display calculator to find the value of a and of b. [3]

Soo Min models the temperature, T, of the water t minutes after it was placed in the room as $T(t) = kc^t + 25$.

(c) Find the value of k and of c. [1]

(d) State a reason why Soo Min's model of the temperature is a better fit for the data than Akira's model. [1]

Question 48

[Maximum mark: 6]

The number of traffic accidents at a road junction is modelled by a Poisson distribution with a mean of 0.76 accidents per week.

(a) Under this model, calculate the probability that

(i) there are at least 2 accidents in a particular week.

(ii) there will be exactly 3 accidents in a particular 4-week period. [4]

The local traffic authority wishes to determine the probability that, in an 8-week period, fewer than 2 accidents occur in a week on exactly 5 occasions. It assumes that the weekly occurrence of accidents is independent of the week in which these occur.

(b) State the appropriate model that the traffic authority should use to determine this probability. [2]

Question 49

[Maximum mark: 7]

Gustav plays a game in which he first tosses an unbiased coin and then rolls an unbiased six-sided die.

If the coin shows tails, the score on the die is Gustav's final number of points.

If the coin shows heads, one is added to the score on the die for Gustav's final number of points.

(a) Find the probability that Gustav's final number of points is 7. [2]

(b) Complete the following table. [3]

Final number of points	1	2	3	4	5	6	7
Probability							

(c) Calculate the expected value of Gustav's final number of points. [2]

Question 50

[Maximum mark: 6]

Jerry makes handcrafted chocolates. On average, 1 in 25 of the chocolates that Jerry makes is flawed. Whether or not a chocolate is flawed is independent of all other chocolates.

(a) In a batch of 20 chocolates, chosen at random, find the probability that

(i) two are flawed.

(ii) more than two are flawed. [4]

Jerry sells the perfect chocolates for 50 pesos each and the flawed ones for 15 pesos each.

(b) Calculate the expected number of pesos Jerry makes from selling a batch of 20 randomly selected chocolates. [2]

Question 51

[Maximum mark: 8]

Roma is told by the manufacturers of their phone that, when charging from 0%, the percentage P of charge can be modelled by $P = 100 - 100 \times 2^{-t}$, where t is the number of hours since the phone started charging.

(a) Find the time taken for the amount of charge to go from 0% to 40%, according to this model. [2]

Roma decides to test the model. They charge the phone from 0% and record the charge of the phone for different values of t. The results are shown in the following table.

t **hours**	1	2	3	4
P **percent**	48	74	86	91

(b) Find the sum of the square residuals (SS_{res}) when using $P = 100 - 100 \times 2^{-t}$ to fit this data. [4]

Roma now uses a quadratic function to model the charge in the phone. They calculate that, when using this model, the sum of the square residuals for these four points is 2.45.

(c) State one reason why Roma might prefer to use

(i) the quadratic model.

(ii) $P = 100 - 100 \times 2^{-t}$. [2]

Question 52

[Maximum mark: 6]

The formula $F = 1.8C + 32$ is used to convert a temperature in degrees Celsius, C, to degrees Fahrenheit, F.

(a) (i) Find a formula for converting a temperature in degrees Fahrenheit to degrees Celsius.

(ii) Find the temperature in degrees Celsius that is recorded as 77 degrees Fahrenheit. [3]

Over one year, the mean daily temperature in Mexico City was calculated to be 17 degrees Celsius with a standard deviation of 9 degrees Celsius.

(b) For the same year, find in degrees Fahrenheit

(i) the mean daily temperature in Mexico City.

(ii) the standard deviation of the daily temperature in Mexico City. [3]

Question 53

[Maximum marks: 7]

After taking a mathematics test, Fatima wonders how many more marks she would have achieved if she had spent an extra 1.5 hours studying.

To find out, she randomly selects five students from her class who took the same test and asks them how many hours (t) they spent studying for the test and the marks (m) they achieved. Their responses are shown in the following table.

Hours, *t*	0	1.2	1.6	2.5	4
Marks, *m*	45	54	61	72	86

Fatima believes there might be a linear relationship between the time spent studying and the results obtained.

(a) (i) Find the Pearson's product moment correlation coefficient, r, for this data.

(ii) Find the least squares regression line of m on t for this data. [4]

(b) According to her model, find how many more marks Fatima would have achieved if she spent an extra 1.5 hours studying. [2]

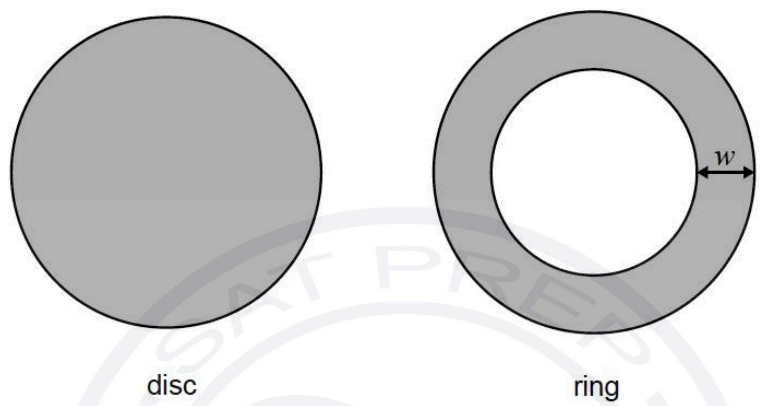
(c) State one reason why the value obtained in part (b) might not be valid. [1]

Question 54

[Maximum mark: 8]

A metal ring, of width w, is made by first cutting a circular disc and then cutting a circular hole exactly in the centre of this disc. This is shown in the diagram.

diagram not to scale



A machine produces many of these rings each day.

The diameters of the discs are normally distributed with mean 30 cm and standard deviation 0.8 cm.

(a) Find the distribution of the radii of the discs. [3]

The radii of the **holes** are normally distributed with mean 12 cm and standard deviation 0.25 cm.

(b) Calculate the probability that the width of a randomly chosen ring is less than 2.5 cm. [5]

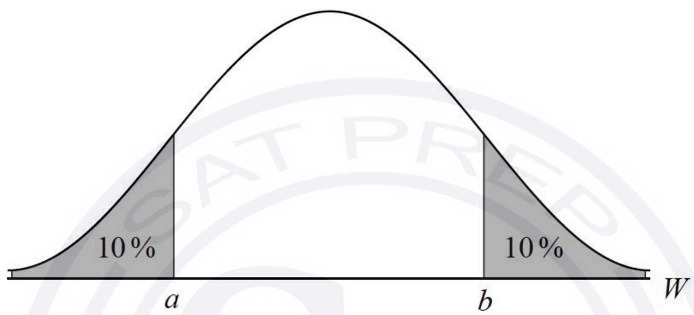
Question 55

[Maximum mark: 7]

The mass, W, of Manx cats is normally distributed with a mean of $4.5\,\text{kg}$ and a standard deviation of $0.4\,\text{kg}$.

(a) A Manx cat is selected at random. Calculate the probability this cat's mass is more than $3.5\,\text{kg}$. [2]

The following curve represents this distribution. It is known that $\mathrm{P}(W < a) = 0.1$ and $\mathrm{P}(W > b) = 0.1$.



(b) Find the value of

(i) a

(ii) b. [3]

(c) Two Manx cats are selected at random from a large population. Find the probability that they both have a mass less than $3.5\,\text{kg}$. [2]