

Subject – Math AI(Standard 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.

- (a) Find the probability that on any given day Mr Burke chooses a female student to answer a question. [1]

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

- (b) Find the probability he will choose a female student 8 times. [2]
(c) Find the probability he will choose a male student at most 9 times. [3]

Question 2

[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 3

[Maximum mark: 6]

Ms Calhoun measures the heights of students in her mathematics class. She is interested to see if the mean height of male students, μ_1 , is the same as the mean height of female students, μ_2 . The information is recorded in the table.

Male height (cm)	150	148	143	152	151	149	147	
Female height (cm)	148	152	154	147	146	153	152	150

At the 10% level of significance, a t -test was used to compare the means of the two groups. The data is assumed to be normally distributed and the standard deviations are equal between the two groups.

- (a) (i) State the null hypothesis. [2]
(ii) State the alternative hypothesis. [2]
- (b) Calculate the p -value for this test. [2]
- (c) State, giving a reason, whether Ms Calhoun should accept the null hypothesis. [2]

Question 4

[Maximum mark: 5]

As part of a study into healthy lifestyles, Jing visited Surrey Hills University. Jing recorded a person's position in the university and how frequently they ate a salad. Results are shown in the table.

	Salad meals per week			
	0	1–2	3–4	>4
Students	45	26	18	6
Professors	15	8	5	12
Staff and Administration	16	13	10	6

Jing conducted a χ^2 test for independence at a 5% level of significance.

- (a) State the null hypothesis. [1]
- (b) Calculate the p -value for this test. [2]
- (c) State, giving a reason, whether the null hypothesis should be accepted. [2]

Question 5

[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

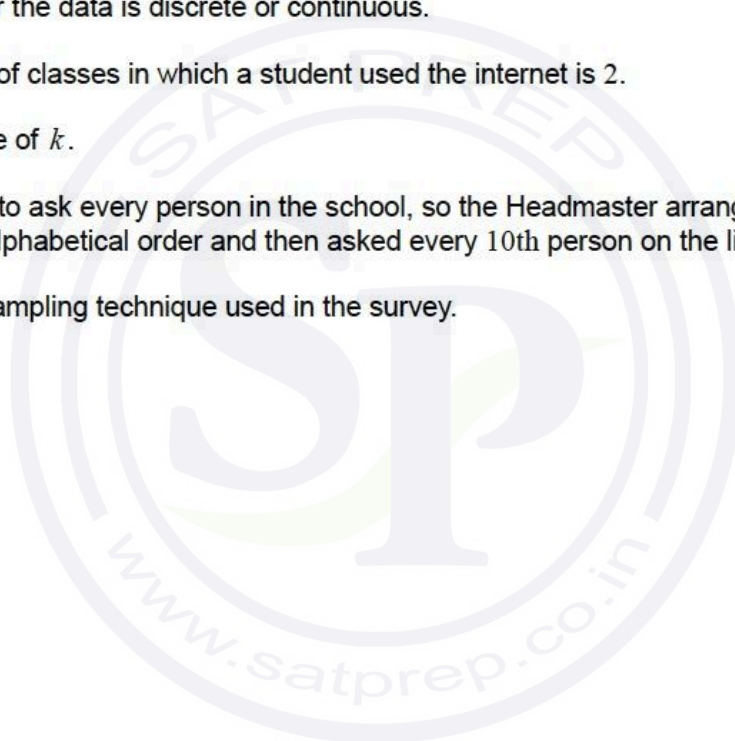
(a) State whether the data is discrete or continuous. [1]

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]

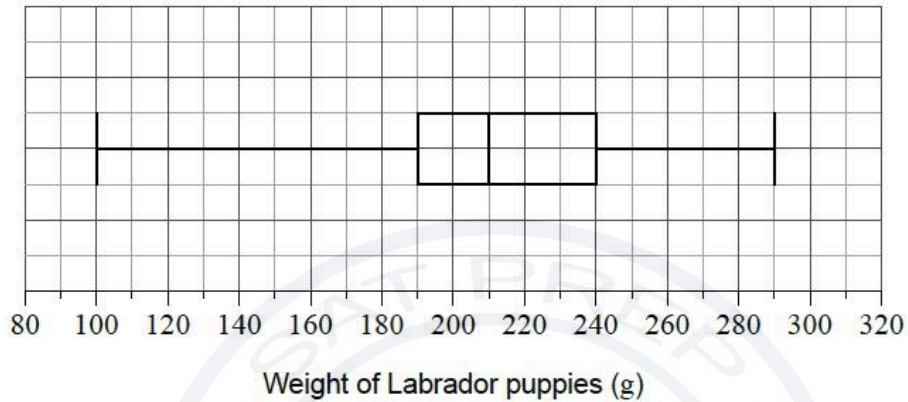


Question 6

[Maximum mark: 6]

Palvinder breeds Labrador puppies at his farm. Over many years he recorded the weight (g) of the puppies.

The data is illustrated in the following box and whisker diagram.



- (a) Write down the median weight of the puppies. [1]
- (b) Write down the upper quartile. [1]
- (c) Find the interquartile range. [2]
- The weights of these Labrador puppies are normally distributed.
- (d) Find the weight of the heaviest possible puppy that is not an outlier. [2]

Question 7

[Maximum mark: 6]

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 and the hypotheses are

H_0 : The data satisfies the model.

H_1 : The data does not satisfy the model.

- (a) Find an estimate for how many copies the vendor expects to sell each day. [1]
- (b) (i) Write down the degrees of freedom for this test.
- (ii) Write down the conclusion to the test. Give a reason for your answer. [5]

Question 8

[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

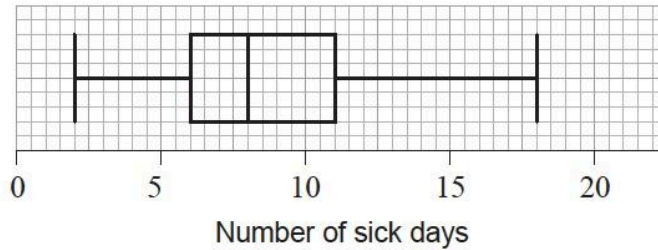
A t -test is to be performed at the 5% significance level.

- (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 9

[Maximum mark: 5]

The number of sick days taken by each employee in a company during a year was recorded. The data was organized in a box and whisker diagram as shown below:



(a) For this data, write down

- (i) the minimum number of sick days taken during the year.
- (ii) the lower quartile.
- (iii) the median.

[3]

Paul claims that this box and whisker diagram can be used to infer that the percentage of employees who took fewer than six sick days is smaller than the percentage of employees who took more than eleven sick days.

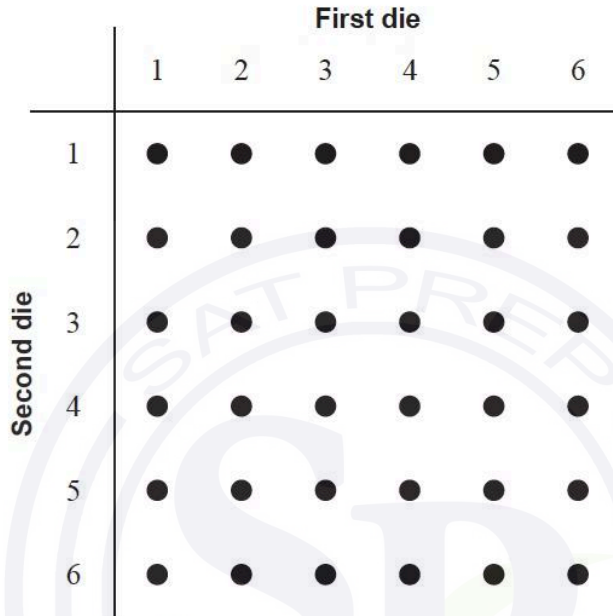
(b) State whether Paul is correct. Justify your answer.

[2]

Question 10

[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 11

[Maximum mark: 5]

Arriane has geese on her farm. She claims the mean weight of eggs from her black geese is less than the mean weight of eggs from her white geese.

She recorded the weights of eggs, in grams, from a random selection of geese. The data is shown in the table.

Weights of eggs from black geese	136	134	142	141	128	126
Weights of eggs from white geese	135	138	141	140	136	134

In order to test her claim, Arriane performs a t -test at a 10% level of significance. It is assumed that the weights of eggs are normally distributed and the samples have equal variances.

- (a) State, in words, the null hypothesis. [1]
- (b) Calculate the p -value for this test. [2]
- (c) State whether the result of the test supports Arriane's claim. Justify your reasoning. [2]

Question 12

[Maximum mark: 4]

Deb used a thermometer to record the maximum daily temperature over ten consecutive days. Her results, in degrees Celsius ($^{\circ}\text{C}$), are shown below.

14, 15, 14, 11, 10, 9, 14, 15, 16, 13

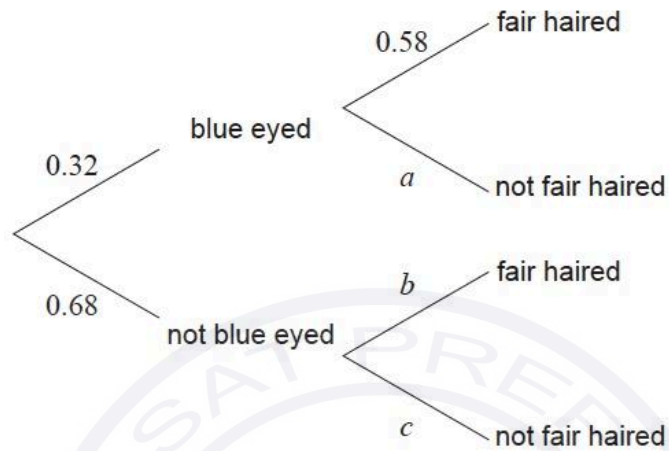
For this data set, find the value of

- (a) the mode. [1]
- (b) the mean. [2]
- (c) the standard deviation. [1]

Question 13

[Maximum mark: 5]

In a city, 32% of people have blue eyes. If someone has blue eyes, the probability that they also have fair hair is 58%. This information is represented in the following tree diagram.



- (a) Write down the value of a . [1]
- (b) Find an expression, in terms of b , for the probability of a person not having blue eyes **and** having fair hair. [1]

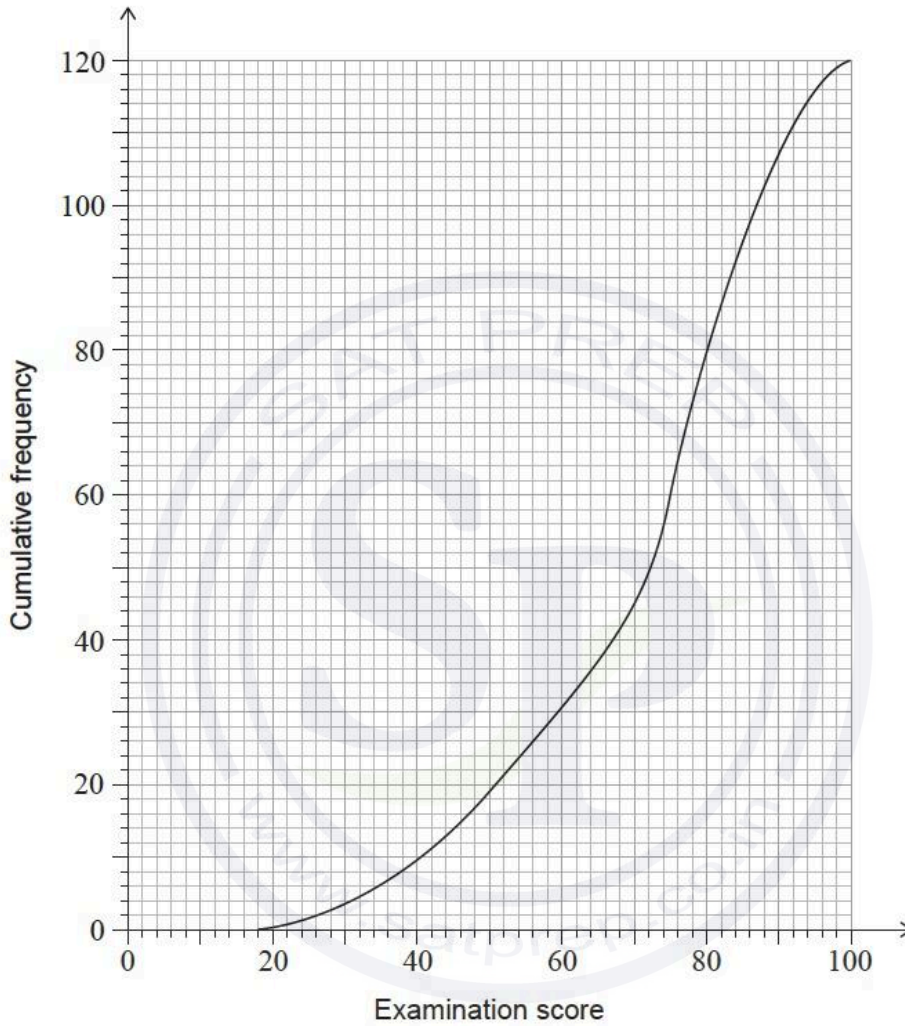
It is known that 41% of people in this city have fair hair.

- (c) Calculate the value of
- (i) b . [3]
- (ii) c .

Question 14

[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]

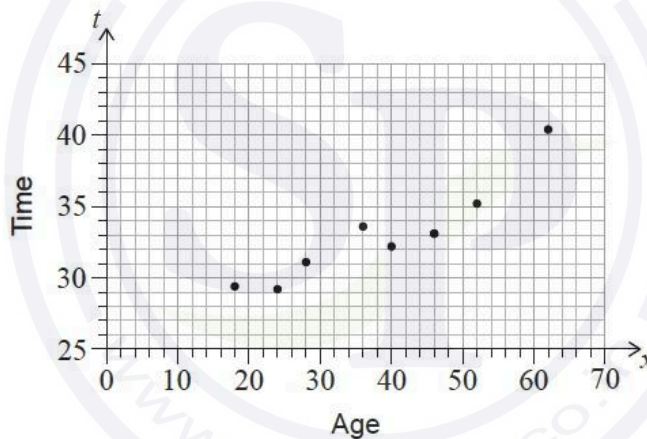
Question 15

[Maximum mark: 6]

Eduardo believes that there is a linear relationship between the age of a male runner and the time it takes them to run 5000 metres.

To test this, he recorded the age, x years, and the time, t minutes, for eight males in a single 5000m race. His results are presented in the following table and scatter diagram.

x, years	18	24	28	36	40	46	52	62
t, minutes	29.4	29.2	31.1	33.6	32.2	33.1	35.2	40.4



(a) For this data, find the value of the Pearson's product-moment correlation coefficient, r . [2]

Eduardo looked in a sports science text book. He found that the following information about r was appropriate for athletic performance.

Value of $ r $	Description of the correlation
$0 \leq r < 0.4$	weak
$0.4 \leq r < 0.8$	moderate
$0.8 \leq r \leq 1$	strong

(b) Comment on your answer to part (a), using the information that Eduardo found. [1]

(c) Write down the equation of the regression line of t on x , in the form $t = ax + b$. [1]

Question 18

[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 19

[Maximum mark: 5]

The masses of Fuji apples are normally distributed with a mean of 163 g and a standard deviation of 6.83 g.

When Fuji apples are picked, they are classified as small, medium, large or extra large depending on their mass. Large apples have a mass of between 172 g and 183 g.

- (a) Determine the probability that a Fuji apple selected at random will be a large apple. [2]

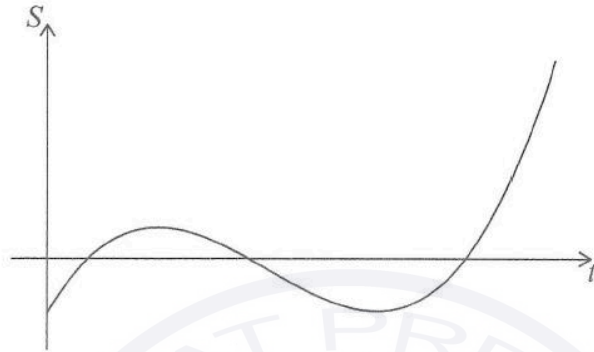
Approximately 68% of Fuji apples have a mass within the medium-sized category, which is between k and 172 g.

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

Question 20

[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 21

[Maximum mark: 6]

A study was conducted to investigate whether the mean reaction time of drivers who are talking on mobile phones is the same as the mean reaction time of drivers who are talking to passengers in the vehicle. Two independent groups were randomly selected for the study.

To gather data, each driver was put in a car simulator and asked to either talk on a mobile phone or talk to a passenger. Each driver was instructed to apply the brakes as soon as they saw a red light appear in front of the car. The reaction times of the drivers, in seconds, were recorded, as shown in the following table.

Talking on mobile phone	Talking to passenger
0.69	0.67
0.87	0.86
0.98	0.60
1.04	0.81
0.79	0.76
0.87	0.71
0.71	0.74

At the 10% level of significance, a t -test was used to compare the mean reaction times of the two groups. Each data set is assumed to be normally distributed, and the population variances are assumed to be the same.

Let μ_1 and μ_2 be the population means for the two groups. The null hypothesis for this test is $H_0: \mu_1 - \mu_2 = 0$.

- (a) State the alternative hypothesis. [1]
- (b) Calculate the p -value for this test. [2]
- (c) (i) State the conclusion of the test. Justify your answer.
- (ii) State what your conclusion means in context. [3]

Question 22

[Maximum mark: 5]

A college runs a mathematics course in the morning. Scores for a test from this class are shown below.

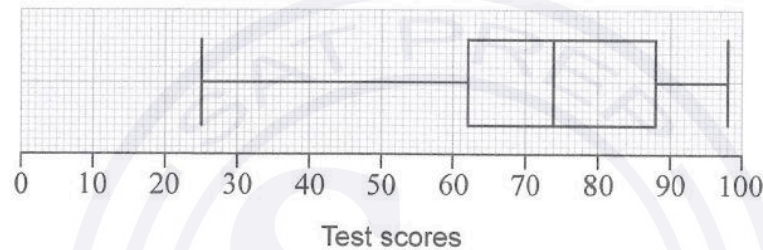
25 33 51 62 63 63 70 74 79 79 81 88 90 90 98

For these data, the lower quartile is 62 and the upper quartile is 88.

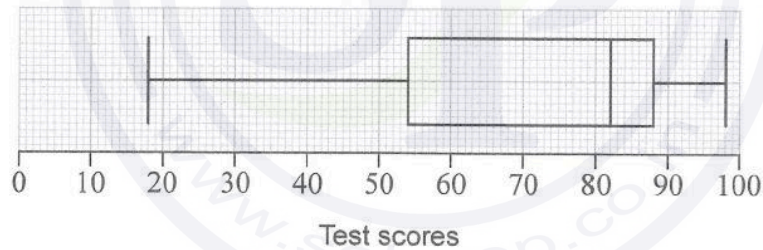
(a) Show that the test score of 25 would not be considered an outlier.

[3]

The box and whisker diagram showing these scores is given below.



Another mathematics class is run by the college during the evening. A box and whisker diagram showing the scores from this class for the same test is given below.



A researcher reviews the box and whisker diagrams and believes that the evening class performed better than the morning class.

(b) With reference to the box and whisker diagrams, state one aspect that may support the researcher's opinion and one aspect that may counter it.

[2]

Question 23

[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 24

[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 25

[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 26

[Maximum mark: 5]

Roy is a member of a motorsport club and regularly drives around the Port Campbell racetrack.

The times he takes to complete a lap are normally distributed with mean 59 seconds and standard deviation 3 seconds.

- (a) Find the probability that Roy completes a lap in less than 55 seconds. [2]

Roy will complete a 20 lap race. It is expected that 8.6 of the laps will take more than t seconds.

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

Question 27

[Maximum mark: 5]

Manny and Annabelle, mathematics teachers at Burnham High School, give their students the same examination. A random sample of the examination scores were collected from each of their classes.

Examination scores from Manny's class	76	77	82	84	88	90	91	98
Examination scores from Annabelle's class	68	79	81	89	91	92	92	95

Annabelle uses these scores to conduct a two-tailed t -test to compare the means of the two classes, at the 5% level of significance. It is assumed the examination scores for both classes have the same variance and are normally distributed.

The null hypothesis is $\mu_1 = \mu_2$, where μ_1 is the mean examination score from Manny's class and μ_2 is the mean examination score from Annabelle's class.

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

(b) Find the p -value for this test. Give your answer correct to five decimal places. [2]

Annabelle concludes there is insufficient evidence to reject the null hypothesis.

(c) State whether Annabelle's conclusion is correct. Give a reason for your answer. [2]

Question 28

[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 29

[Maximum mark: 7]

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 km h^{-1} .

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

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

Speeding tickets are issued to all drivers travelling at a speed greater than 72 km h^{-1} .

(b) Find the probability that a randomly selected driver who passes the speed camera receives a speeding ticket. [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 km h^{-1} .

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

Question 30

[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 31

[Maximum mark: 6]

The decathlon is a competition where athletes compete in ten events. Two of those events are long jump and high jump. In both events, a greater distance means a better ranking.

The table shows results for these two events at the World Championships.

Athlete's Country	Event		Rank	
	Long Jump (m)	High Jump (m)	Long Jump Rank	High Jump Rank
Germany	7.64	2.11	1	
France	7.52	2.08	2	
Estonia	7.49	1.84	3	
Canada	7.44	2.02	4	
Netherlands	7.33	2.05	5	
Ukraine	7.28	2.02	6	
Algeria	7.22	1.90	7	
Austria	7.11	1.87	8	
Grenada	6.98	1.99	9	
Japan	6.64	1.96	10	

The Spearman's rank correlation coefficient is used to determine if there is a linear correlation between an athlete's ranking in long jump and their ranking in high jump.

- (a) Complete the table to show the athletes' rankings in high jump. [2]
- (b) Find the value of the Spearman's rank correlation coefficient r_s . [2]

The following guide is used by the coach to determine the strength of the correlation between the ranks for long jump and high jump.

$ r_s $	Strength
0.000 to 0.199	Very weak
0.200 to 0.399	Weak
0.400 to 0.599	Moderate
0.600 to 0.799	Strong
0.800 to 1.000	Very strong

- (c) State the strength of the correlation between the rankings as indicated by the table and interpret this in the context of the question. [2]

Question 32

[Maximum mark: 5]

In a game, balls are thrown to hit a target. The random variable X is the number of times the target is hit in five attempts. The probability distribution for X is shown in the following table.

x	0	1	2	3	4	5
$P(X=x)$	0.15	0.2	k	0.16	$2k$	0.25

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

The player has a chance to win money based on how many times they hit the target.

The gain for the player, in \$, is shown in the following table, where a negative gain means that the player loses money.

x	0	1	2	3	4	5
Player's gain (\$)	-4	-3	-1	0	1	4

- (b) Determine whether this game is fair. Justify your answer. [3]

Question 33

[Maximum mark: 6]

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

A seed from this mango tree is chosen at random.

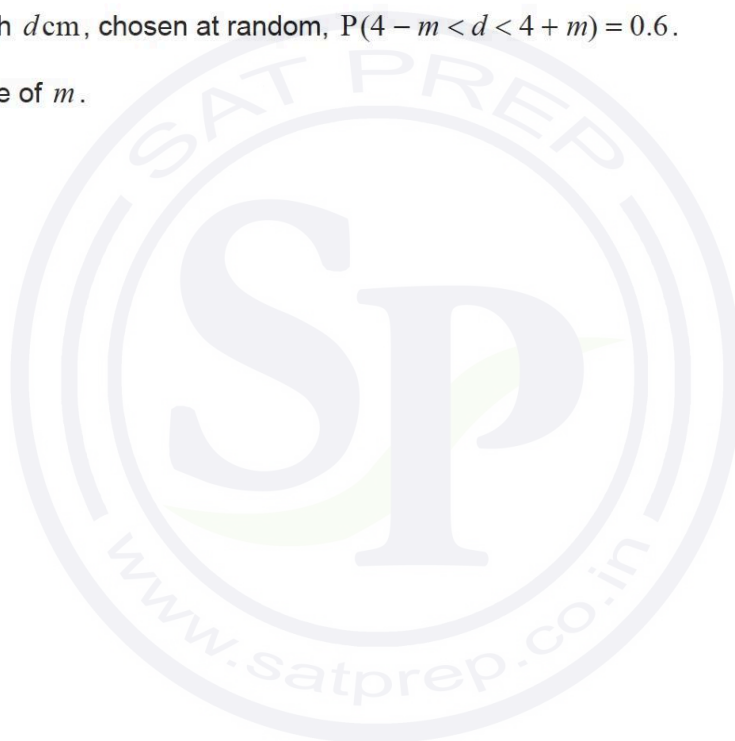
- (a) Calculate the probability that the length of the seed is less than 3.7 cm. [2]

It is known that 30% of the seeds have a length greater than k cm.

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

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

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



Question 34

[Maximum mark: 6]

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

An item of food is chosen at random from these 500.

- (a) Find the probability that its quality is not perfect, given that it is from breakfast. [2]

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.

- (b) Find the χ^2 statistic. [2]
- (c) State, with justification, the conclusion for this test. [2]

Question 35

[Maximum mark: 4]

The following frequency distribution table shows the test grades for a group of students.

Grade	1	2	3	4	5	6	7
Frequency	1	4	7	9	p	9	4

For this distribution, the mean grade is 4.5.

- (a) Write down the total number of students in terms of p . [1]
- (b) Calculate the value of p . [3]

Question 36

[Maximum mark: 6]

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 100m. 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]

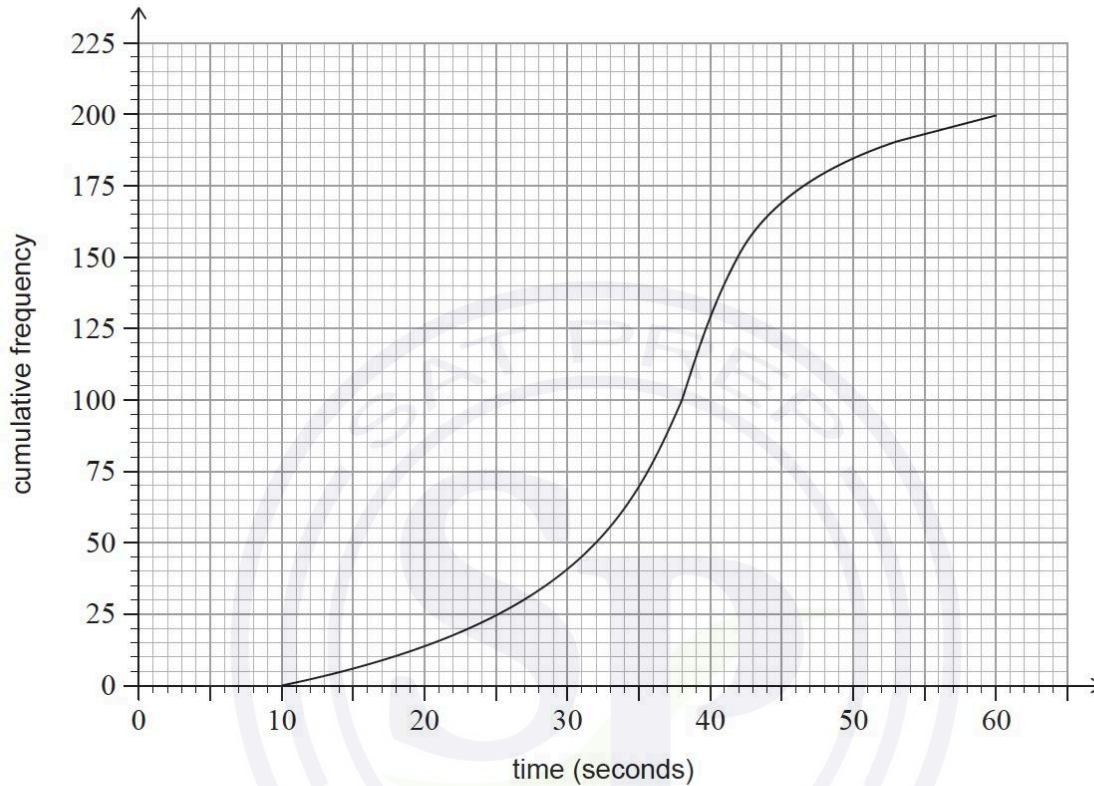
Athlete	A	B	C	D	E	F	G	H
Age rank			3					
Time rank							1	

- (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]

Question 37

[Maximum mark: 7]

In a school, 200 students solved a problem in a mathematics competition. Their times to solve the problem were recorded and the following cumulative frequency graph was produced.



(a) Use the graph to find

- (i) the median time;
- (ii) the lower quartile;
- (iii) the upper quartile;
- (iv) the interquartile range.

[4]

Cedric took 14 seconds to solve the problem.

(b) Determine whether Cedric's time is an outlier.

[3]

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

Thurston believes that artists with a higher number of social media followers sell more albums in the first week. He carries out a hypothesis test using a 10% significance level with the following null hypothesis:

H_0 : In the population, there is no monotonic relationship between the number of social media followers and the number of albums sold in the first week.

- (c) Write down Thurston's alternative hypothesis. [1]

The critical value of r_s for this test is 0.643.

- (d) State the conclusion of the hypothesis test, giving a reason. [2]

Question 39

[Maximum mark: 6]

Nicole works at a local school 5 days each week. She drives an old car to work that has a 72% probability of starting on any given morning. The probability of the car starting on a given morning is independent of it starting on any other morning.

- (a) Find the probability that Nicole's car starts on exactly three mornings in a particular 5 day workweek. [2]

Nicole walks to work on mornings when her car does not start and it is **not** raining. Nicole takes the bus to work on mornings when her car does not start and it is raining.

Where Nicole lives, there is a 42% probability of rain on any given morning, independent of any other morning. The probability of Nicole's car starting is independent of the weather.

- (b) Find the probability that Nicole will **not** have to take the bus in a particular workweek. [4]

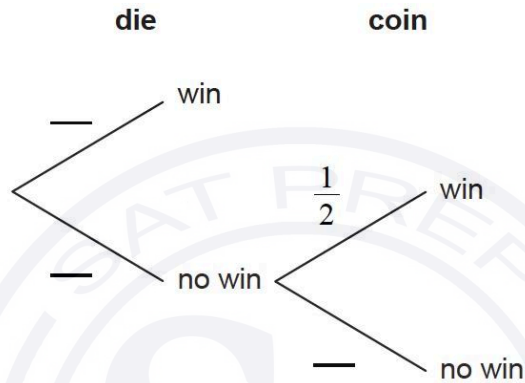
Question 40

[Maximum mark: 7]

Rita is playing a game. In the game, she must roll a fair six-sided die. If she gets a five or six then she wins a prize. If not, then she has another chance but this time she must flip a fair coin which will result in the coin landing on heads or tails. If the coin lands on heads, then Rita wins a prize.

- (a) Complete the tree diagram by writing in the three missing probabilities.

[2]



- (b) Find the probability that Rita does **not** win a prize.

[2]

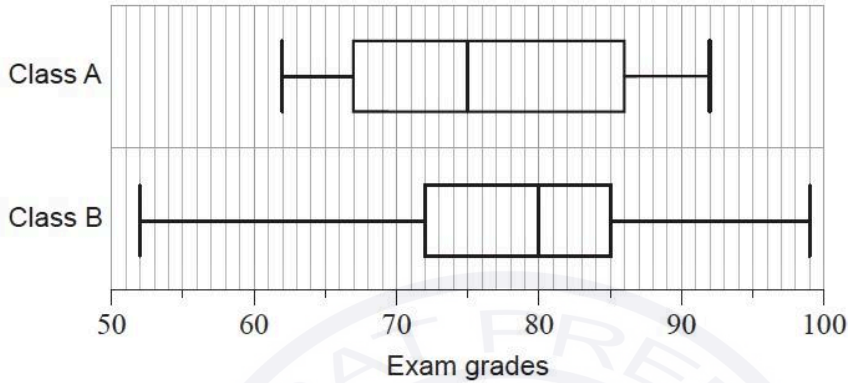
- (c) Given that Rita won a prize, find the probability that she got a five or six when she rolled the die.

[3]

Question 41

[Maximum mark: 8]

Mrs Whitehouse is a chemistry teacher. After grading her final exams, she creates the following box and whisker diagram to compare the grades of her two classes.



- (a) Identify which **two** of the following statements **must** be true according to the box and whisker diagram. Indicate your choices by placing tick marks in the second column of the following table.

[2]

Statement	True (✓)
The data for Class A is normally distributed.	
A higher percentage of students in Class A received a grade less than 70 on the exam, than in Class B.	
More students in Class B received a grade greater than 90 on the exam than in Class A.	
The interquartile range for Class B is less than the interquartile range for Class A.	

At the end of the year, Mrs Whitehouse surveyed a random sample of students from each of her two large classes to determine how satisfied they were with her teaching.

Each student independently selected a value from 1 to 10, with 1 meaning that they were not satisfied at all and 10 meaning that they were very satisfied.

Her collected data from the student surveys is shown.

Class A	7	5	3	4	3	8	6	5		
Class B	6	9	8	10	1	9	10	9	8	3

Mrs Whitehouse believes that there was no difference in the general satisfaction between the two classes. She assumes that the data is drawn from a population that can be modelled by a normal distribution and proposes to conduct a t -test at the 5% significance level.

- (b) Write down the null and alternative hypotheses for her test. [2]
- (c) Find the p -value for her test. [2]
- (d) Write down the conclusion to the test. Give a reason for your answer. [2]

Question 42

[Maximum mark: 7]

Billy is a keen walker who keeps a record of his performance. The following table shows the time, in minutes, it takes him to walk one kilometre up hills with different gradients. The gradient of each hill is constant.

Gradient G (%)	0	4	10	15	20
Time T (min.)	6.85	8.42	11.20	14.49	17.88

- (a) (i) Find the equation of the regression line of T on G .
- (ii) Describe the correlation between T and G with reference to the value of r , the Pearson's product-moment correlation coefficient. [4]

On Sunday, Billy intends to walk up a hill with a gradient of 13%.

- (b) Estimate the time it will take Billy to walk one kilometre up the hill. [2]

This morning, Billy walked one kilometre up a hill, and it took 22 minutes.

- (c) Explain why it would be inappropriate to use the equation found in part (a) to estimate the gradient of this hill. [1]

Question 43

2. [Maximum mark: 6]

Thurston believes that more popular musical artists sell more albums.

He begins to investigate this belief by randomly selecting eight musical artists and collecting data on the number of followers each of the artists has on a particular social media platform. He then collects data on the number of albums each artist sold in the first week after releasing an album. His data is shown in **Table 1**.

Table 1

	Artist 1	Artist 2	Artist 3	Artist 4	Artist 5	Artist 6	Artist 7	Artist 8
Number of social media followers (in thousands)	11 500	12 400	1300	2300	674	49 500	315	94 400
Number of albums sold in first week (in thousands)	123	62.4	17.4	94.9	52.5	27	21.6	595.5

Thurston decides to calculate the Spearman's rank correlation coefficient.

(a) Complete the table of ranks shown in **Table 2**.

[1]

Table 2

	Artist 1	Artist 2	Artist 3	Artist 4	Artist 5	Artist 6	Artist 7	Artist 8
Rank – social media followers	4	3	6	5	7	2	8	1
Rank – albums sold in first week								1

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

Thurston believes that artists with a higher number of social media followers sell more albums in the first week. He carries out a hypothesis test using a 10% significance level with the following null hypothesis:

H_0 : In the population, there is no monotonic relationship between the number of social media followers and the number of albums sold in the first week.

- (c) Write down Thurston's alternative hypothesis. [1]

The critical value of r_s for this test is 0.643.

- (d) State the conclusion of the hypothesis test, giving a reason. [2]

Question 44

[Maximum mark: 6]

Toktam works at a local bakery 5 days each week. She drives an old car to work that has a 65% probability of starting on any given morning. The probability of the car starting on a given morning is independent of it starting on any other morning.

- (a) Find the probability that Toktam's car starts on exactly two mornings in a particular 5 day workweek. [2]

Toktam walks to work on mornings when her car does not start and it is **not** raining. Toktam takes a taxi to work on mornings when her car does not start and it is raining.

Where Toktam lives, there is a 45% probability of rain on any given morning, independent of any other morning. The probability of Toktam's car starting is independent of the weather.

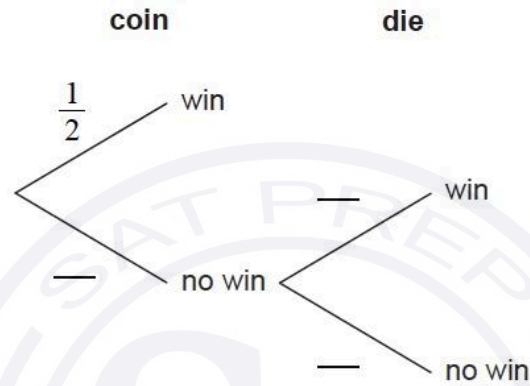
- (b) Find the probability that Toktam will **not** have to take a taxi in a particular workweek. [4]

Question 45

[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 die 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]

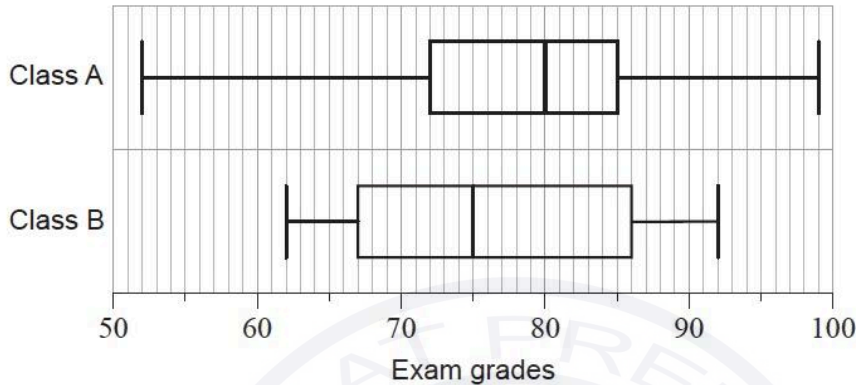


- (b) Find the probability that Michèle does **not** win a prize. [2]
- (c) Given that Michèle won a prize, find the probability that the coin landed on heads. [3]

Question 46

[Maximum mark: 8]

Mr Kelly is a history teacher. After grading his final exams, he creates the following box and whisker diagram to compare the grades of his two classes.



- (a) Identify which **two** of the following statements **must** be true according to the box and whisker diagram. Indicate your choices by placing tick marks in the second column of the following table.

[2]

Statement	True (✓)
A higher percentage of students in Class B received a grade less than 70 on the exam, than in Class A.	
The data for Class B is normally distributed.	
More students in Class A received a grade greater than 90 on the exam than in Class B.	
The interquartile range for Class A is less than the interquartile range for Class B.	

At the end of the year, Mr Kelly surveyed a random sample of students from each of his two large classes to determine how satisfied they were with his teaching.

Each student independently selected a value from 1 to 10, with 1 meaning that they were not satisfied at all and 10 meaning that they were very satisfied.

His collected data from the student surveys is shown.

Class A	6	9	8	10	1	9	10	9	8	4
Class B	7	5	3	4	3	8	6	7		

Mr Kelly believes that there was no difference in the general satisfaction between the two classes. He assumes that the data is drawn from a population that can be modelled by a normal distribution and proposes to conduct a t -test at the 5% significance level.

- (b) Write down the null and alternative hypotheses for his test. [2]
- (c) Find the p -value for his test. [2]
- (d) Write down the conclusion to the test. Give a reason for your answer. [2]

Question 47

[Maximum mark: 7]

Joel is a keen cyclist who keeps a record of his performance. The following table shows the time, in minutes, it takes him to ride one kilometre on hills with different gradients. The gradient of each hill is constant.

Gradient G (%)	0	4	10	15	20
Time T (min.)	2.11	5.39	10.56	13.20	18.58

- (a) (i) Find the equation of the regression line of T on G .
- (ii) Describe the correlation between T and G with reference to the value of r , the Pearson's product-moment correlation coefficient. [4]

On Saturday, Joel intends to ride a hill with a gradient of 17%.

- (b) Estimate the time it will take Joel to ride one kilometre up the hill. [2]

This morning, Joel rode one kilometre up a hill, and it took 22 minutes.

- (c) Explain why it would be inappropriate to use the equation found in part (a) to estimate the gradient of this hill. [1]

Question 48

[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 49

[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 50

[Maximum mark: 5]

Sunita sorts 300 peppers into sizes of small, medium or large. Some peppers are red, some are green, and some are yellow.

The following table shows her results.

	Small	Medium	Large
Red	18	31	46
Green	26	32	21
Yellow	34	66	26

Sunita wants to test, at the 5% significance level, whether the size of the peppers is independent of the colour.

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

The critical value for this test is 9.49.

- (b) (i) Calculate χ^2_{calc} . [4]
(ii) State a conclusion to the test. Give a reason for your answer.

Question 51

[Maximum mark: 7]

The prices, in dollars, of 10 different garden chairs are:

79 139 255 99 50 209 229 193 69 49

- (a) Find the range of the prices of the 10 chairs. [2]
(b) Use your graphic display calculator to find
(i) the mean price of the chairs.
(ii) the standard deviation of the price of the chairs. [3]

In a sale, the price of each of the 10 garden chairs is reduced by \$20.

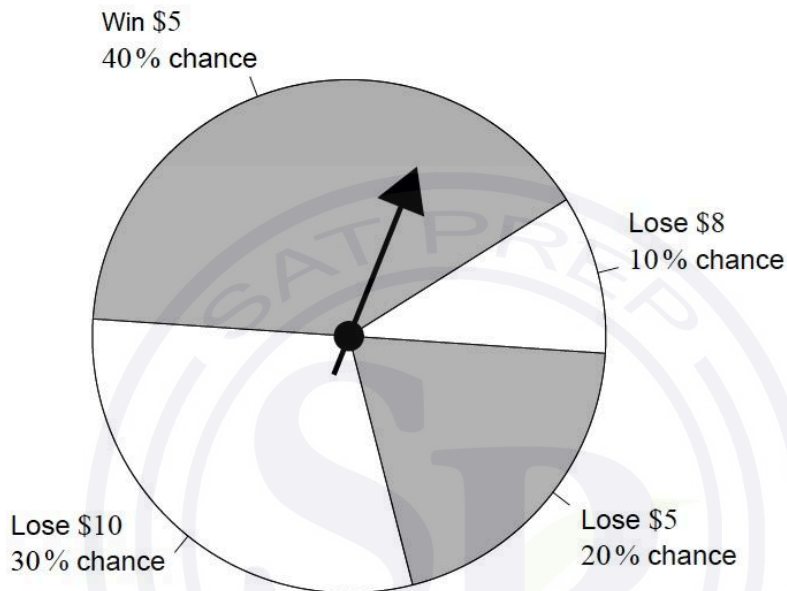
- (c) Write down
(i) the new mean.
(ii) the new standard deviation. [2]

Question 52

[Maximum mark: 8]

Zac raises funds for a library by running a game where players spin a needle. The final position of the needle results in an outcome where a player wins or loses money. The outcomes, with associated probabilities, are shown in the following diagram.

diagram not to scale



Let X represent the amount that a player of this game wins.

- (a) (i) Find the expected value of X .
- (ii) Interpret your answer to part (a)(i).

[3]

To encourage a person to keep playing this game, Zac increases the winning prize for the second game they play from \$5 to \$6. For each successive game they play, the winning prize continues to increase by \$1.

Emily plays k games. The k th game is fair.

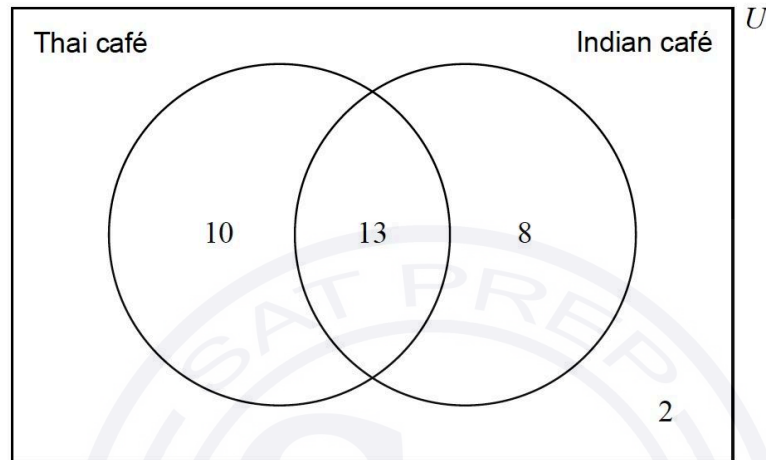
- (b) (i) Find the value of k .
- (ii) Explain why Zac expects to raise money from the games Emily plays.

[5]

Question 53

[Maximum mark: 6]

A teacher surveys their students to find out if they have eaten at the local Thai and Indian cafés. The results of the survey are shown in the following Venn diagram.



(a) Write down the number of students surveyed. [1]

(b) Write down the number of students who have not eaten at the Indian café. [1]

A student is chosen at random from those surveyed.

(c) Find the probability this student has eaten at both the Thai café and the Indian café. [1]

Let T be the event: a student has eaten at the Thai café.

Let I be the event: a student has eaten at the Indian café.

(d) Find $P(T \cup I)$. [1]

(e) State whether the events T and I are mutually exclusive. Justify your answer. [2]

Question 54

[Maximum mark: 7]

The following data show the heights, in metres, of six players in a basketball team.

1.67	1.60	1.68	2.31	2.31	2.19
------	------	------	------	------	------

(a) For these six players, find

(i) the mean height.

(ii) the median height.

(iii) the modal height.

(iv) the range of the heights.

[6]

A new player, Gheorghe, joins the team. Their height is measured as 1.98 metres to the nearest centimetre.

(b) Write down the shortest possible height of Gheorghe.

[1]

Question 55

[Maximum mark: 7]

A fair game is played where points are scored as follows:

- A win scores w points.
- A draw scores 0 points.
- A loss scores -5 points.

Let X be the number of points scored during a game. The probability distribution is shown.

x	w	0	-5
$P(X=x)$	0.25	0.4	p

(a) Find the value of p .

[2]

The game is played 60 times.

(b) Find the expected number of losses.

[2]

(c) Calculate the value of w , given that the game is fair.

[3]

Question 56

[Maximum mark: 7]

The records at a driving school show that 55% of students pass their driving test on the first attempt.

A group of 20 students take their driving test for the first time.

As part of its quality control, the driving school uses the model $X \sim B(20, 0.55)$, where X is the number of students who pass the driving test.

- (a) Calculate the
- (i) mean of X
 - (ii) variance of X . [2]
- (b) Find the probability that
- (i) exactly 14 students pass the test
 - (ii) fewer than 5 students pass the test. [4]
- (c) State one assumption that the driving school makes in using this model. [1]

Question 57

[Maximum mark: 7]

The records at a driving school show that 55% of students pass their driving test on the first attempt.

A group of 20 students take their driving test for the first time.

As part of its quality control, the driving school uses the model $X \sim B(20, 0.55)$, where X is the number of students who pass the driving test.

- (a) Calculate the
- (i) mean of X
 - (ii) variance of X . [2]
- (b) Find the probability that
- (i) exactly 14 students pass the test
 - (ii) fewer than 5 students pass the test. [4]
- (c) State one assumption that the driving school makes in using this model. [1]

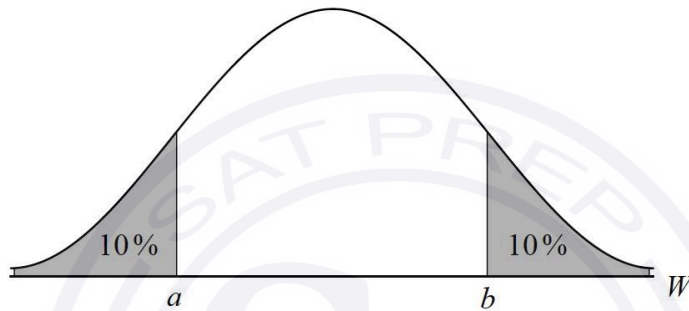
Question 58

[Maximum mark: 7]

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

- (a) A Manx cat is selected at random. Calculate the probability this cat's mass is more than 3.5 kg. [2]

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



- (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 kg. [2]

Question 59

[Maximum mark: 7]

The scores on a test, out of 7 points, for 220 students are shown in the following table.

Score	1	2	3	4	5	6	7
Frequency	11	24	26	35	61	43	20

- (a) Find the mean **and** standard deviation of these test scores. Give your answers correct to four significant figures. [3]

- (b) The scores are multiplied by ten. Write down the new mean and standard deviation. [2]

- (c) After the scores have been multiplied by ten, 30 points are added to each of them. Write down the new mean and standard deviation. [2]

Question 60

[Maximum mark: 7]

A fair game is played where points are scored as follows:

- A win scores w points.
- A draw scores 0 points.
- A loss scores -7 points.

Let X be the number of points scored during a game. The probability distribution is shown.

x	w	0	-7
$P(X=x)$	0.35	0.4	p

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

The game is played 60 times.

(b) Find the expected number of losses. [2]

(c) Calculate the value of w , given that the game is fair. [3]

Question 61

[Maximum mark: 7]

The records at a driving school show that 60% of students pass their driving test on the first attempt.

A group of 30 students take their driving test for the first time.

As part of its quality control, the driving school uses the model $X \sim B(30, 0.60)$, where X is the number of students who pass the driving test.

(a) Calculate the
(i) mean of X
(ii) variance of X . [2]

(b) Find the probability that
(i) exactly 21 students pass the test
(ii) fewer than 12 students pass the test. [4]

(c) State one assumption that the driving school makes in using this model. [1]

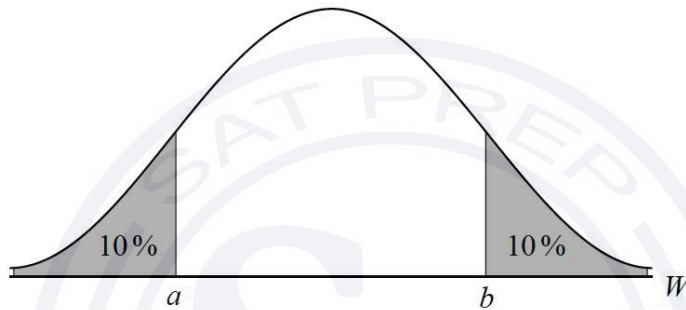
Question 62

[Maximum mark: 7]

The mass, W , of Scottish Terriers (a breed of dog) is normally distributed with a mean of 8.9 kg and a standard deviation of 1.2 kg.

- (a) A Scottish Terrier is selected at random. Calculate the probability this dog's mass is more than 7.2 kg. [2]

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



- (b) Find the value of

(i) a

(ii) b . [3]

- (c) Two Scottish Terriers are selected at random from a large population. Find the probability that both dogs have a mass less than 7.2 kg. [2]

Question 63

[Maximum mark: 7]

The scores on a test, out of 7 points, for 240 students are shown in the following table.

Score	1	2	3	4	5	6	7
Frequency	11	29	31	34	65	47	23

- (a) Find the mean **and** standard deviation of these test scores. Give your answers correct to four significant figures. [3]

- (b) The scores are multiplied by ten. Write down the new mean and standard deviation. [2]

- (c) After the scores have been multiplied by ten, 30 points are added to each of them. Write down the new mean and standard deviation. [2]