Subject – Math AI(Standard Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2022
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
$\mathbf{P}(X=x)$	$\frac{1}{18}$	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.

(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.

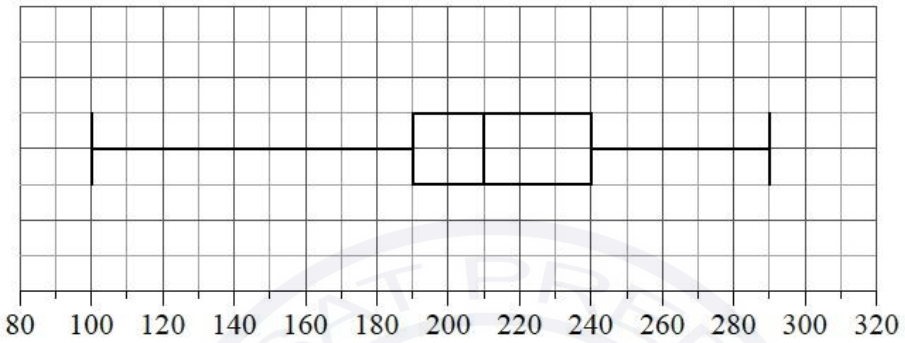
(c) Identify the sampling technique used in the survey. [1]

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.



Weight of Labrador puppies (g)

(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]

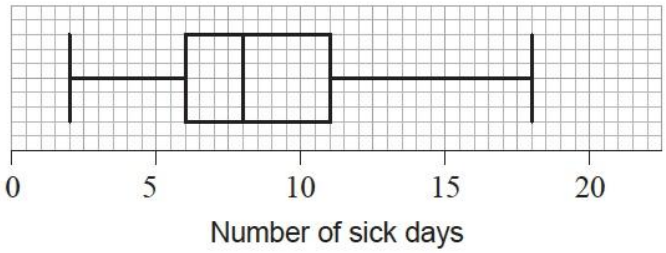
(b) Find the p-value for this test. [2]

(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]

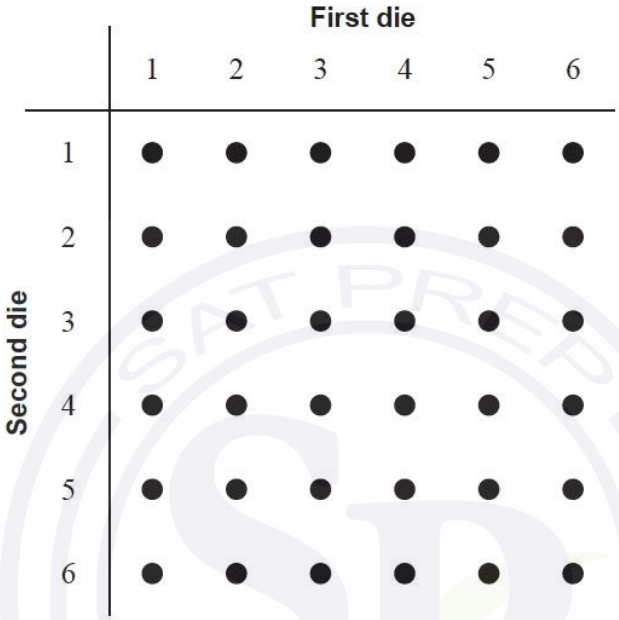
Paul claims that this box and whisker diagram can be used to infer that the percentage of employees who took fewer than six sick days is smaller than the percentage of employees who took more than eleven sick days.

(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
$\mathbf{P}(T=t)$						

(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 ($^{\circ}$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]

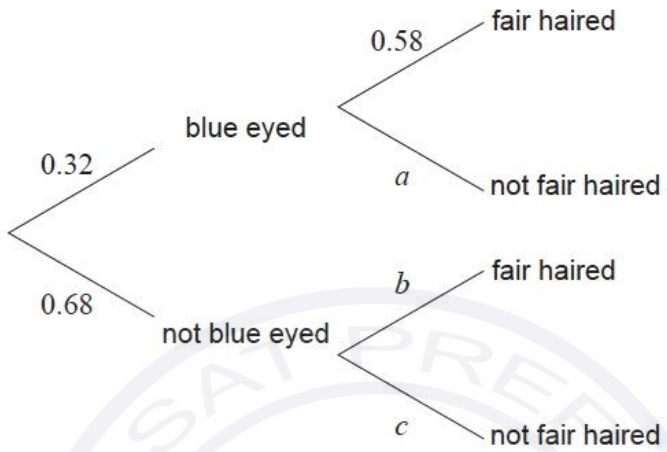
(b) the mean. [2]

(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

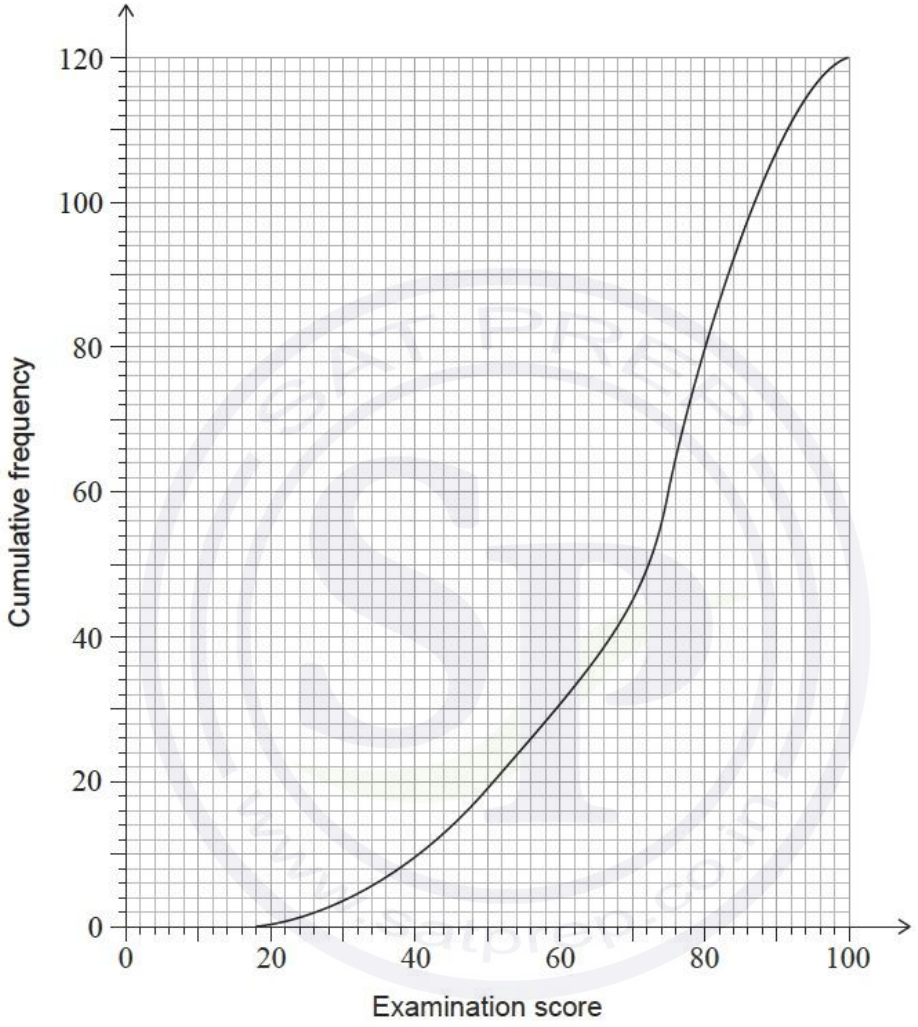
(i) b.

(ii) c. [3]

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]

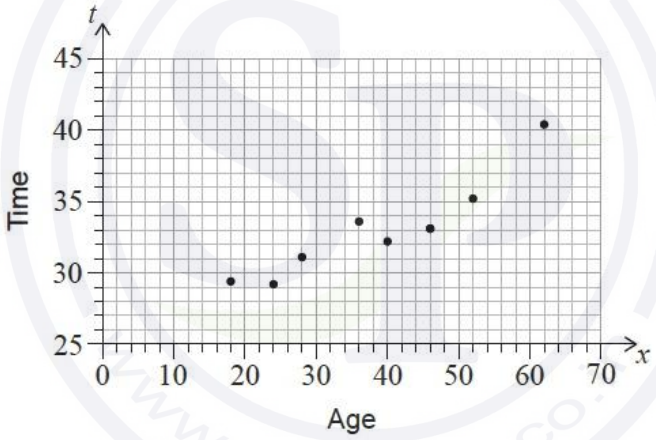
Question 15

[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 $5000\,\text{m}$ 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 $\lvert r \rvert$	**Description of the correlation**
$0 \leq \lvert r \rvert < 0.4$	weak
$0.4 \leq \lvert r \rvert < 0.8$	moderate
$0.8 \leq \lvert r \rvert \leq 1$	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]

A 57-year-old male also ran in the $5000\,\text{m}$ race.

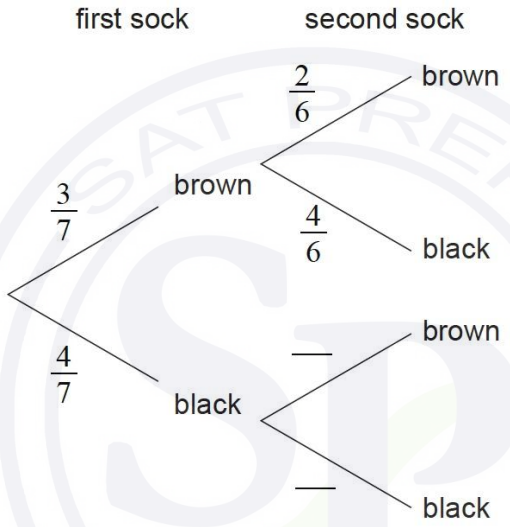
(d) Use the equation of the regression line to estimate the time he took to complete the $5000\,\text{m}$ race. [2]

Question 16

[Maximum mark: 6]

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. [1]



(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]

Question 17

[Maximum mark: 6]

A factory produces bags of sugar with a labelled weight of $500\,\text{g}$. The weights of the bags are normally distributed with a mean of $500\,\text{g}$ and a standard deviation of $3\,\text{g}$.

(a) Write down the percentage of bags that weigh more than $500\,\text{g}$. [1]

A bag that weighs less than $495\,\text{g}$ is rejected by the factory for being underweight.

(b) Find the probability that a randomly chosen bag is rejected for being underweight. [2]

A bag that weighs more than k grams is rejected by the factory for being overweight.
The factory rejects $2\,\%$ of bags for being overweight.

(c) Find the value of k. [3]

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 $163\,\text{g}$ and a standard deviation of $6.83\,\text{g}$.

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 $172\,\text{g}$ and $183\,\text{g}$.

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

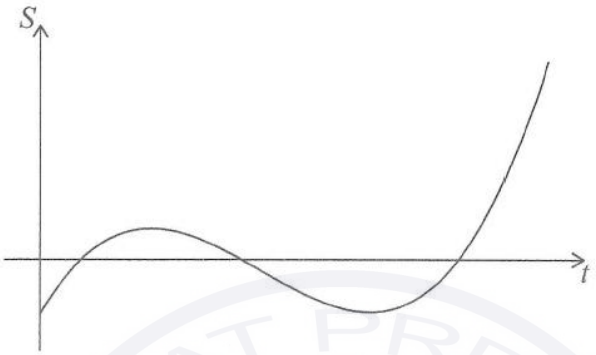
Approximately 68% of Fuji apples have a mass within the medium-sized category, which is between k and $172\,\text{g}$.

(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]

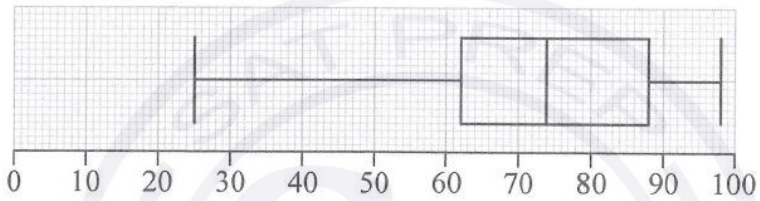
A college runs a mathematics course in the morning. Scores for a test from this class are shown below.

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

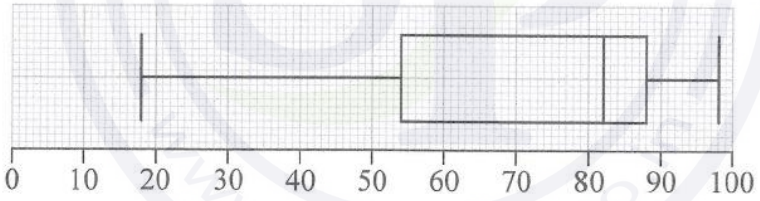
For these data, the lower quartile is 62 and the upper quartile is 88.

(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]