

APOLLO

Statistics and probability

SL

Mathematics: applications and interpretation

First assessment 2021

112 questions | questions

2. [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]

[illegible]

Turn over

6. [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

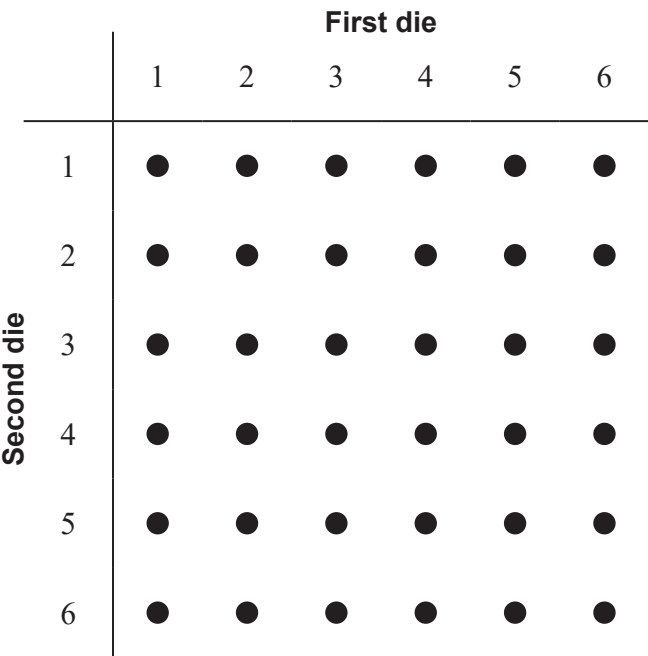
In order to test her claim, Arriane performs a t -test at a 10% level of significance. It is assumed that the weights of eggs are normally distributed and the samples have equal variances.

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

[illegible]

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]

(This question continues on the following page).



(Question 10 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

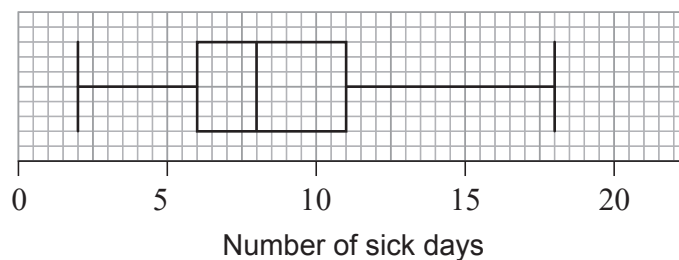
.....



Turn over

5. [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

- the minimum number of sick days taken during the year.
- the lower quartile.
- 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]

[illegible]

Turn over

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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

11. [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]

[illegible]

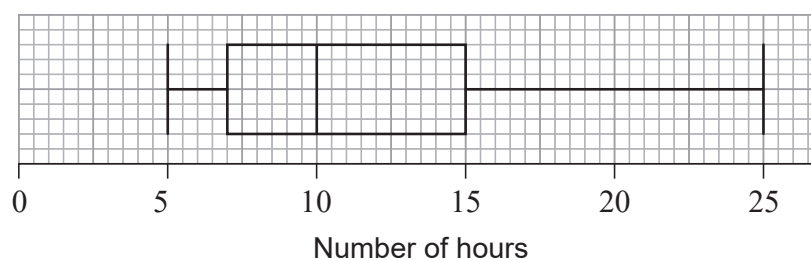
Turn over

1. [Maximum mark: 18]

As part of his mathematics exploration about classic books, Jason investigated the time taken by students in his school to read the book *The Old Man and the Sea*. He collected his data by stopping and asking students in the school corridor, until he reached his target of 10 students from **each** of the literature classes in his school.

- (a) State which of the two sampling methods, systematic or quota, Jason has used. [1]

Jason constructed the following box and whisker diagram to show the number of hours students in the sample took to read this book.



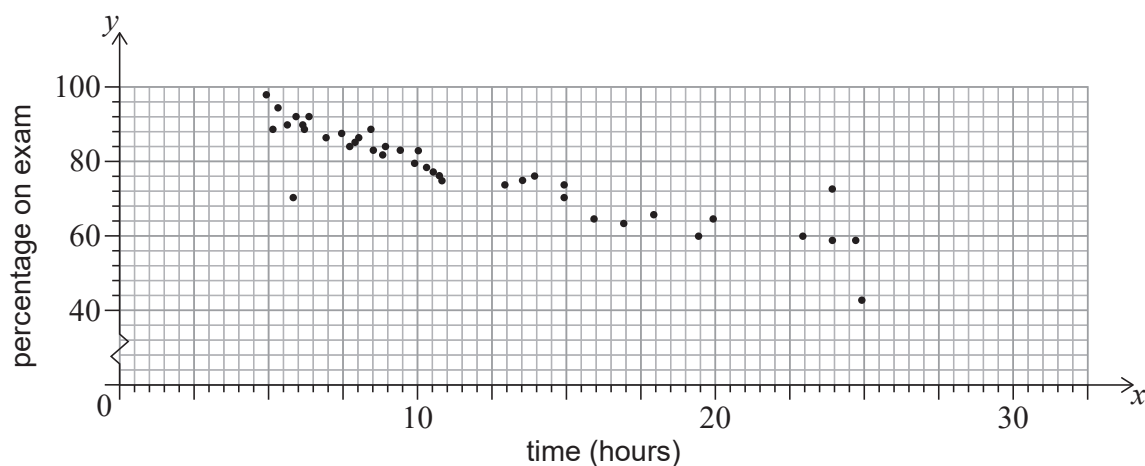
- (b) Write down the median time to read the book. [1]

- (c) Calculate the interquartile range. [2]

Mackenzie, a member of the sample, took 25 hours to read the novel. Jason believes Mackenzie's time is not an outlier.

- (d) Determine whether Jason is correct. Support your reasoning. [4]

For each student interviewed, Jason recorded the time taken to read *The Old Man and the Sea* (x), measured in hours, and paired this with their percentage score on the final exam (y). These data are represented on the scatter diagram.



- (e) Describe the correlation. [1]

(This question continues on the following page)

(Question 1 continued)

Jason correctly calculates the equation of the regression line y on x for these students to be

$$y = -1.54x + 98.8.$$

He uses the equation to estimate the percentage score on the final exam for a student who read the book in 1.5 hours.

- (f) Find the percentage score calculated by Jason. [2]
- (g) State whether it is valid to use the regression line y on x for Jason's estimate. Give a reason for your answer. [2]

Jason found a website that rated the 'top 50' classic books. He randomly chose eight of these classic books and recorded the number of pages. For example, Book H is rated 44th and has 281 pages. These data are shown in the table.

Book	A	B	C	D	E	F	G	H
Number of pages (n)	4215	863	585	1225	366	209	624	281
Top 50 rating (t)	1	2	5	7	13	22	40	44

Jason intends to analyse the data using Spearman's rank correlation coefficient, r_s .

- (h) Copy and complete the information in the following table. [2]

Book	A	B	C	D	E	F	G	H
Rank – Number of pages	1							
Rank – Top 50 Rating	1							

- (i) (i) Calculate the value of r_s . [3]
- (ii) Interpret your result.

Turn over

4. [Maximum mark: 13]

The stopping distances for bicycles travelling at 20 km h^{-1} are assumed to follow a normal distribution with mean 6.76 m and standard deviation 0.12 m .

- (a) Under this assumption, find, correct to four decimal places, the probability that a bicycle chosen at random travelling at 20 km h^{-1} manages to stop

(i) in less than 6.5 m .

(ii) in more than 7 m .

[3]

1000 randomly selected bicycles are tested and their stopping distances when travelling at 20 km h^{-1} are measured.

- (b) Find, correct to four significant figures, the expected number of bicycles tested that stop between

(i) 6.5 m and 6.75 m .

(ii) 6.75 m and 7 m .

[3]

The measured stopping distances of the 1000 bicycles are given in the table.

Measured stopping distance	Number of bicycles
Less than 6.5 m	12
Between 6.5 m and 6.75 m	428
Between 6.75 m and 7 m	527
More than 7 m	33

It is decided to perform a χ^2 goodness of fit test at the 5% level of significance to decide whether the stopping distances of bicycles travelling at 20 km h^{-1} can be modelled by a normal distribution with mean 6.76 m and standard deviation 0.12 m .

- (c) State the null and alternative hypotheses.

[2]

- (d) Find the p -value for the test.

[3]

- (e) State the conclusion of the test. Give a reason for your answer.

[2]

Turn over

1. [Maximum mark: 19]

A medical centre is testing patients for a certain disease. This disease occurs in 5 % of the population.

They test every patient who comes to the centre on a particular day.

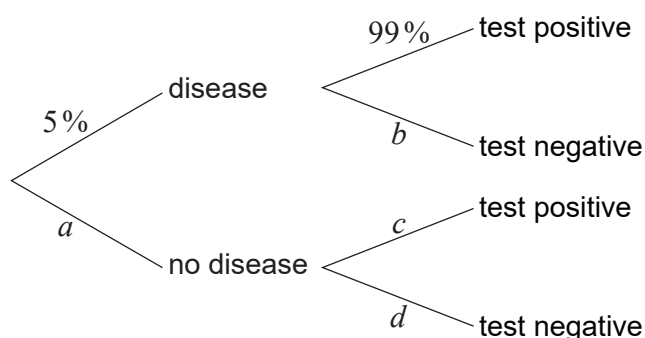
(a) State the sampling method being used.

[1]

It is intended that if a patient has the disease, they test “positive”, and if a patient does not have the disease, they test “negative”.

However, the tests are not perfect, and only 99 % of people who have the disease test positive. Also, 2 % of people who **do not** have the disease test positive.

The tree diagram shows some of this information.



(b) Write down the value of

(i) a .

(ii) b .

(iii) c .

(iv) d .

[4]

(This question continues on the following page)

(Question 1 continued)

- (c) Use the tree diagram to find the probability that a patient selected at random
- (i) will not have the disease and will test positive.
 - (ii) will test negative.
 - (iii) has the disease given that they tested negative. [8]
- (d) The medical centre finds the actual number of positive results in their sample is different than predicted by the tree diagram. Explain why this might be the case. [1]

The staff at the medical centre looked at the care received by all visiting patients on a randomly chosen day. All the patients received at least one of these services: they had medical tests (M), were seen by a nurse (N), or were seen by a doctor (D). It was found that:

- 78 had medical tests,
 - 45 were seen by a nurse;
 - 30 were seen by a doctor;
 - 9 had medical tests and were seen by a doctor and a nurse;
 - 18 had medical tests and were seen by a doctor but were not seen by a nurse;
 - 11 patients were seen by a nurse and had medical tests but were not seen by a doctor;
 - 2 patients were seen by a doctor without being seen by nurse and without having medical tests.
- (e) Draw a Venn diagram to illustrate this information, placing all relevant information on the diagram. [3]
- (f) Find the total number of patients who visited the centre during this day. [2]

Turn over

4. [Maximum mark: 14]

It is known that the weights of male Persian cats are normally distributed with mean 6.1 kg and variance 0.5^2 kg^2 .

(a) Sketch a diagram showing the above information. [2]

(b) Find the proportion of male Persian cats weighing between 5.5 kg and 6.5 kg. [2]

A group of 80 male Persian cats are drawn from this population.

(c) Determine the expected number of cats in this group that have a weight of less than 5.3 kg. [3]

(d) It is found that 12 of the cats weigh more than x kg. Estimate the value of x . [3]

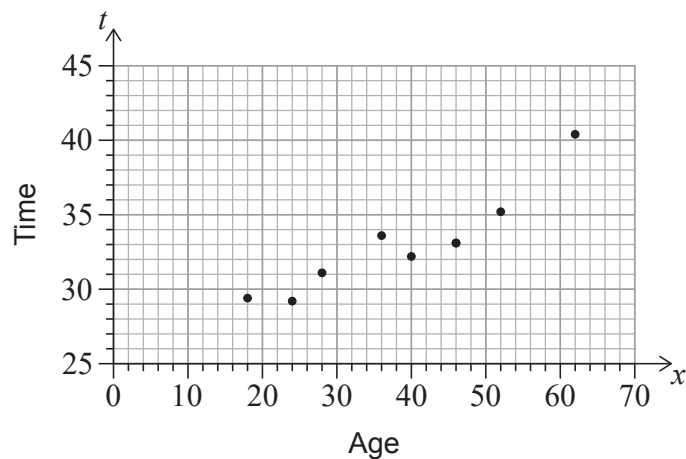
(e) Ten of the cats are chosen at random. Find the probability that exactly one of them weighs over 6.25 kg. [4]

1. [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]

(This question continues on the following page)



A 57-year-old male also ran in the 5000m race.

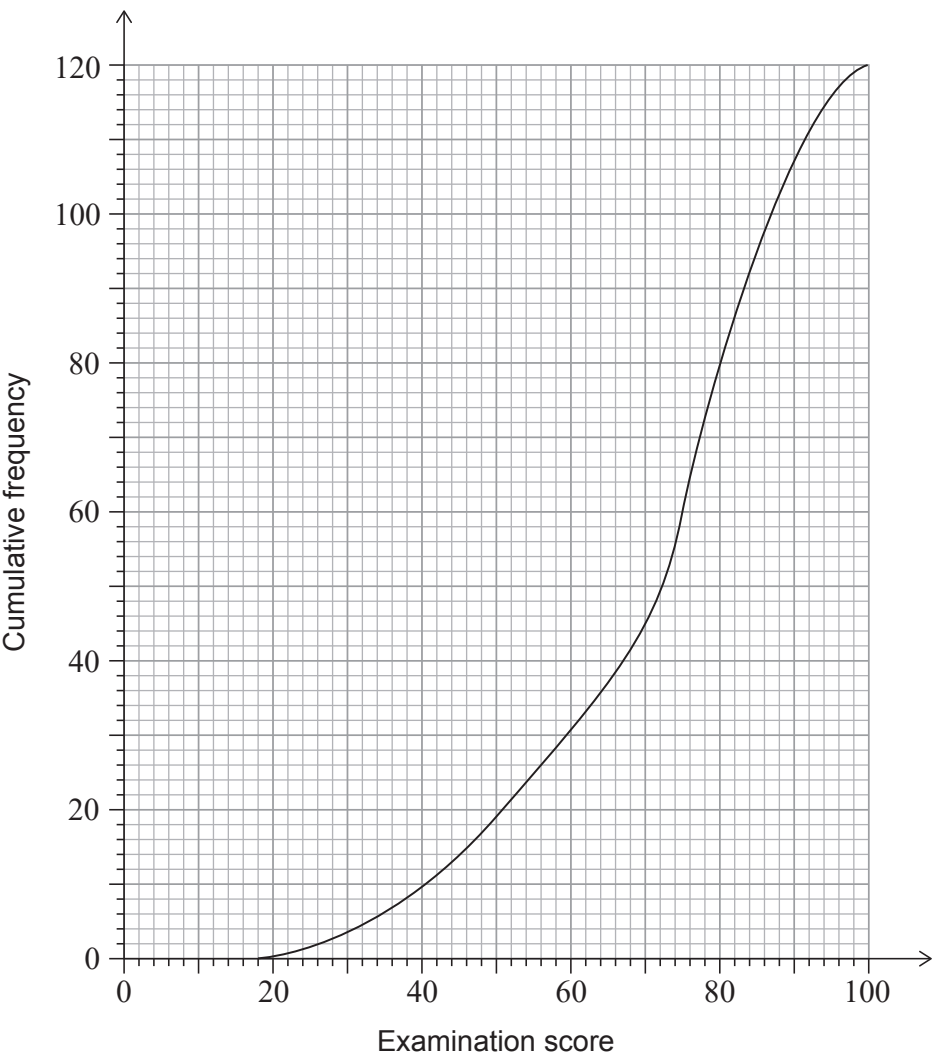
- [2]

This image shows a blank sheet of white paper with horizontal blue ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Apollo | Original source crop | 16

9. [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]

(This question continues on the following page)



(Question 9 continued)

- (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]

[illegible]

Turn over

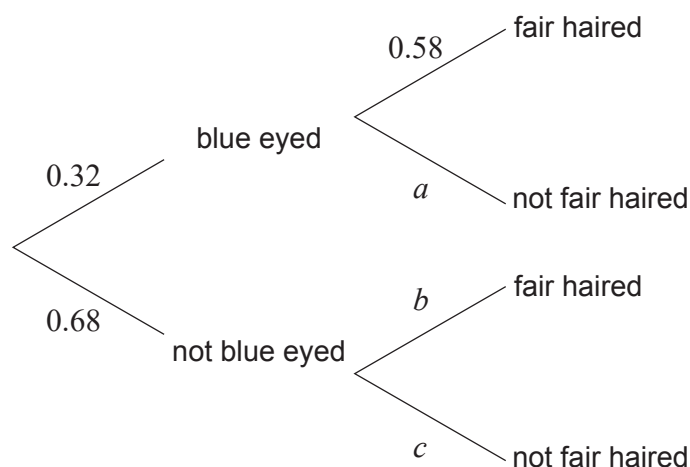
Please **do not** write on this page.

Answers written on this page
will not be marked.



11. [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 .

(This question continues on the following page)



(Question 11 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

1. [Maximum mark: 16]

A group of 1280 students were asked which electronic device they preferred. The results per age group are given in the following table.

	Age			
Preferred device	11–13	14–16	17–18	Total
Laptop	143	160	153	456
Tablet	205	224	131	560
Mobile phone	72	128	64	264
Total	420	512	348	1280

(a) A student from the group is chosen at random. Calculate the probability that the student

- (i) prefers a tablet.
- (ii) is 11–13 years old and prefers a mobile phone.
- (iii) prefers a laptop **given that** they are 17–18 years old.
- (iv) prefers a tablet or is 14–16 years old.

[9]

A χ^2 test for independence was performed on the collected data at the 1 % significance level. The critical value for the test is 13.277.

(b) State the null and alternative hypotheses.

[1]

(c) Write down the number of degrees of freedom.

[1]

(d) (i) Write down the χ^2 test statistic.

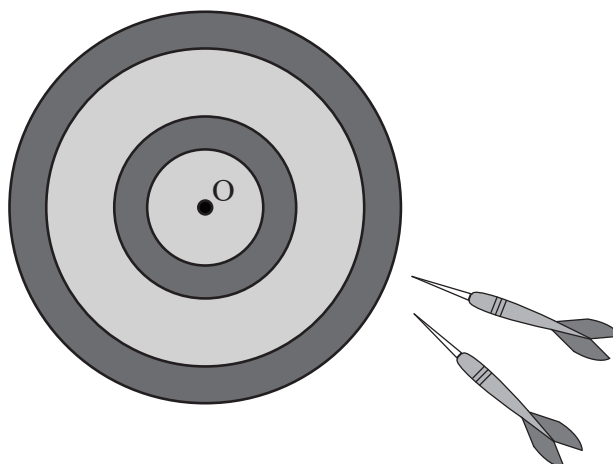
(ii) Write down the p -value.

(iii) State the conclusion for the test in context. Give a reason for your answer.

[5]

5. [Maximum mark: 14]

Arianne plays a game of darts.



The distance that her darts land from the centre, O, of the board can be modelled by a normal distribution with mean 10 cm and standard deviation 3 cm.

(a) Find the probability that

(i) a dart lands less than 13 cm from O.

(ii) a dart lands more than 15 cm from O.

[3]

Each of Arianne's throws is independent of her previous throws.

(b) Find the probability that Arianne throws two consecutive darts that land more than 15 cm from O.

[2]

In a competition a player has three darts to throw on each turn. A point is scored if a player throws **all** three darts to land within a central area around O. When Arianne throws a dart the probability that it lands within this area is 0.8143.

(c) Find the probability that Arianne does **not** score a point on a turn of three darts.

[2]

In the competition Arianne has ten turns, each with three darts.

(d) (i) Find the probability that Arianne scores at least 5 points in the competition.

(ii) Find the probability that Arianne scores at least 5 points and less than 8 points.

(iii) Given that Arianne scores at least 5 points, find the probability that Arianne scores less than 8 points.

[7]

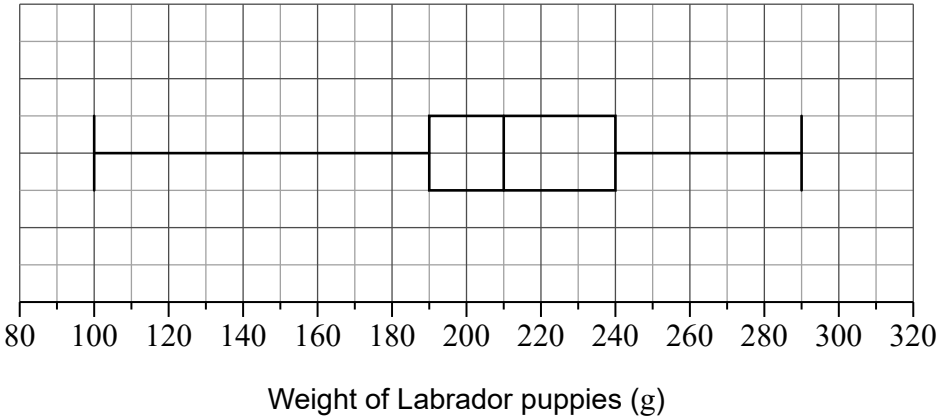
References:

© International Baccalaureate Organization 2021

1. [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]

(This question continues on the following page)



[illegible]

3. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

6. [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.

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

[illegible]

Turn over

9. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

12. [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]



Turn over

13. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



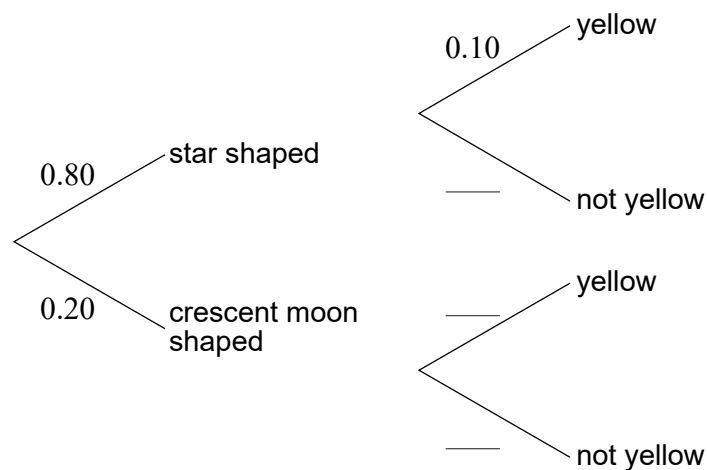
2. [Maximum mark: 14]

Slugworth Candy Company sell a variety pack of colourful, shaped sweets.

The sweets are produced such that 80% are star shaped and 20% are shaped like a crescent moon. It is known that 10% of the stars and 30% of the crescent moons are coloured yellow.

- (a) Using the given information, **copy** and complete the following tree diagram.

[2]



- (b) A sweet is selected at random.

- (i) Find the probability that the sweet is yellow.
- (ii) Given that the sweet is yellow, find the probability it is star shaped.

[4]

(This question continues on the following page)

(Question 2 continued)

According to manufacturer specifications, the colours in each variety pack should be distributed as follows.

Colour	Brown	Red	Green	Orange	Yellow	Purple
Percentage (%)	15	25	20	20	10	10

Mr Slugworth opens a pack of 80 sweets and records the frequency of each colour.

Colour	Brown	Red	Green	Orange	Yellow	Purple
Observed Frequency	10	20	16	18	12	4

To investigate if the sample is consistent with manufacturer specifications, Mr Slugworth conducts a χ^2 goodness of fit test. The test is carried out at a 5 % significance level.

(c) Write down the null hypothesis for this test. [1]

(d) **Copy** and complete the following table in your answer booklet. [2]

Colour	Brown	Red	Green	Orange	Yellow	Purple
Expected Frequency						

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

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

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

Turn over

3. [Maximum mark: 17]

The Malvern Aquatic Center hosted a 3 metre spring board diving event. The judges, Stan and Minsun awarded 8 competitors a score out of 10. The raw data is collated in the following table.

Competitors	A	B	C	D	E	F	G	H
Stan's score (x)	4.1	3	4.3	6	7.1	6	7.5	6
Minsun's score (y)	4.7	4.6	4.8	7.2	7.8	9	9.5	7.2

- (a) (i) Write down the value of the Pearson's product-moment correlation coefficient, r .
 (ii) Using the value of r , interpret the relationship between Stan's score and Minsun's score. [4]
- (b) Write down the equation of the regression line y on x . [2]
- (c) (i) Use your regression equation from part (b) to estimate Minsun's score when Stan awards a perfect 10.
 (ii) State whether this estimate is reliable. Justify your answer. [4]

The Commissioner for the event would like to find the Spearman's rank correlation coefficient.

- (d) **Copy** and complete the information in the following table. [2]

Competitors	A	B	C	D	E	F	G	H
Stan's Rank		8					1	4
Minsun's Rank		8					1	4.5

- (e) (i) Find the value of the Spearman's rank correlation coefficient, r_s .
 (ii) Comment on the result obtained for r_s . [4]

The Commissioner believes Minsun's score for competitor G is too high and so decreases the score from 9.5 to 9.1.

- (f) Explain why the value of the Spearman's rank correlation coefficient r_s does not change. [1]

7. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



8. [Maximum mark: 6]

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

- (a) Write down the percentage of bags that weigh more than 500 g. [1]

A bag that weighs less than 495 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]



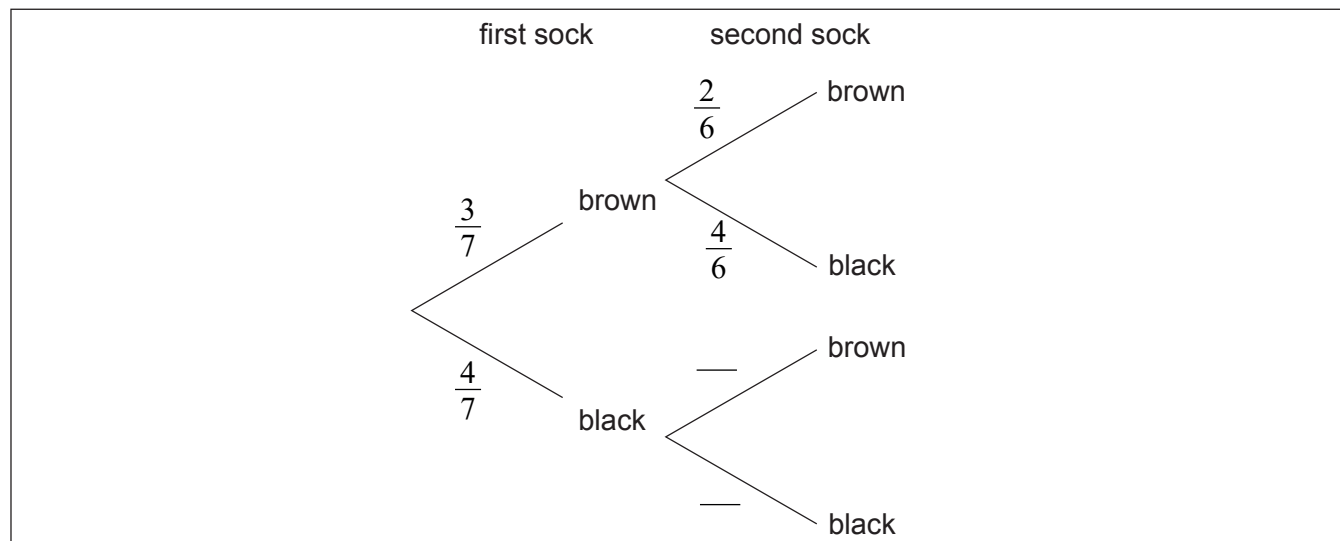
Turn over

10. [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]

[illegible]

Turn over

2. [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]

(This question continues on the following page)





4444-1217

(Question 2 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

5. [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]



Turn over

7. [Maximum mark: 5]

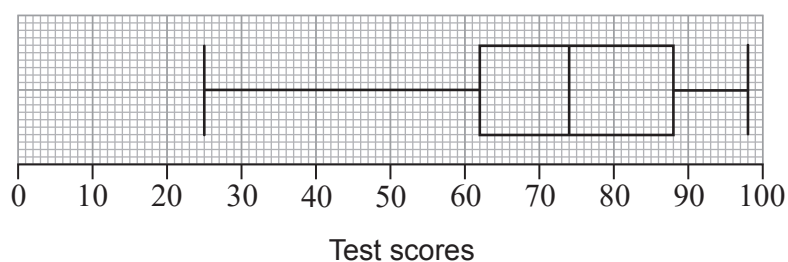
A college runs a mathematics course in the morning. Scores for a test from this class are shown below.

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

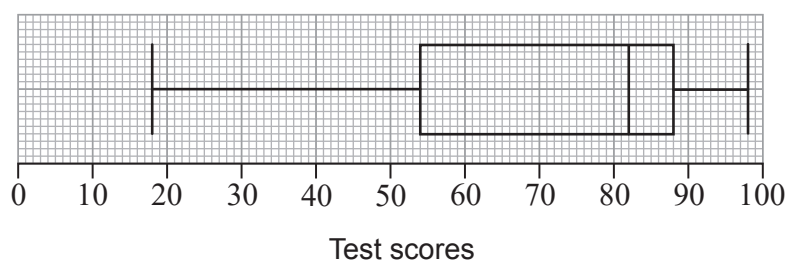
For these data, the lower quartile is 62 and the upper quartile is 88.

(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]

(This question continues on the following page)



(Question 7 continued)

[illegible]

Turn over

8. [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]

(This question continues on the following page)



(Question 8 continued)

[illegible]

Turn over

10. [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]



3. [Maximum mark: 16]

The scores of the eight highest scoring countries in the 2019 Eurovision song contest are shown in the following table.

	Eurovision score
Netherlands	498
Italy	472
Russia	370
Switzerland	364
Sweden	334
Norway	331
North Macedonia	305
Azerbaijan	302

(a) For this data, find

(i) the upper quartile.

(ii) the interquartile range.

[4]

(b) Determine if the Netherlands' score is an outlier for this data. Justify your answer.

[3]

(This question continues on the following page)

Turn over

(Question 3 continued)

Chester is investigating the relationship between the highest-scoring countries' Eurovision score and their population size to determine whether population size can reasonably be used to predict a country's score.

The populations of the countries, to the nearest million, are shown in the table.

	Population (x) (millions)	Eurovision score (y)
Netherlands	17	498
Italy	60	472
Russia	145	370
Switzerland	9	364
Sweden	10	334
Norway	5	331
North Macedonia	2	305
Azerbaijan	10	302

Chester finds that, for this data, the Pearson's product moment correlation coefficient is $r = 0.249$.

- (c) State whether it would be appropriate for Chester to use the equation of a regression line for y on x to predict a country's Eurovision score. Justify your answer.

[2]

Chester then decides to find the Spearman's rank correlation coefficient for this data, and creates a table of ranks.

	Population rank (to the nearest million)	Eurovision score rank
Netherlands	3	1
Italy	2	2
Russia	1	3
Switzerland	a	4
Sweden	b	5
Norway	7	6
North Macedonia	8	7
Azerbaijan	c	8

(This question continues on the following page)

(Question 3 continued)

(d) Write down the value of:

(i) a ,

(ii) b ,

(iii) c .

[3]

(e) (i) Find the value of the Spearman's rank correlation coefficient r_s .

(ii) Interpret the value obtained for r_s .

[3]

(f) When calculating the ranks, Chester incorrectly read the Netherlands' score as 478. Explain why the value of the Spearman's rank correlation r_s does not change despite this error.

[1]

Turn over

5. [Maximum mark: 15]

The aircraft for a particular flight has 72 seats. The airline's records show that historically for this flight only 90% of the people who purchase a ticket arrive to board the flight. They assume this trend will continue and decide to sell extra tickets and hope that no more than 72 passengers will arrive.

The number of passengers that arrive to board this flight is assumed to follow a binomial distribution with a probability of 0.9.

- (a) The airline sells 74 tickets for this flight. Find the probability that more than 72 passengers arrive to board the flight. [3]
- (b) (i) Write down the expected number of passengers who will arrive to board the flight if 72 tickets are sold. [2]
- (ii) Find the maximum number of tickets that could be sold if the expected number of passengers who arrive to board the flight must be less than or equal to 72. [2]

Each passenger pays \$150 for a ticket. If too many passengers arrive, then the airline will give \$300 in compensation to each passenger that cannot board.

- (c) Find, to the nearest integer, the expected increase or decrease in the money made by the airline if they decide to sell 74 tickets rather than 72. [8]

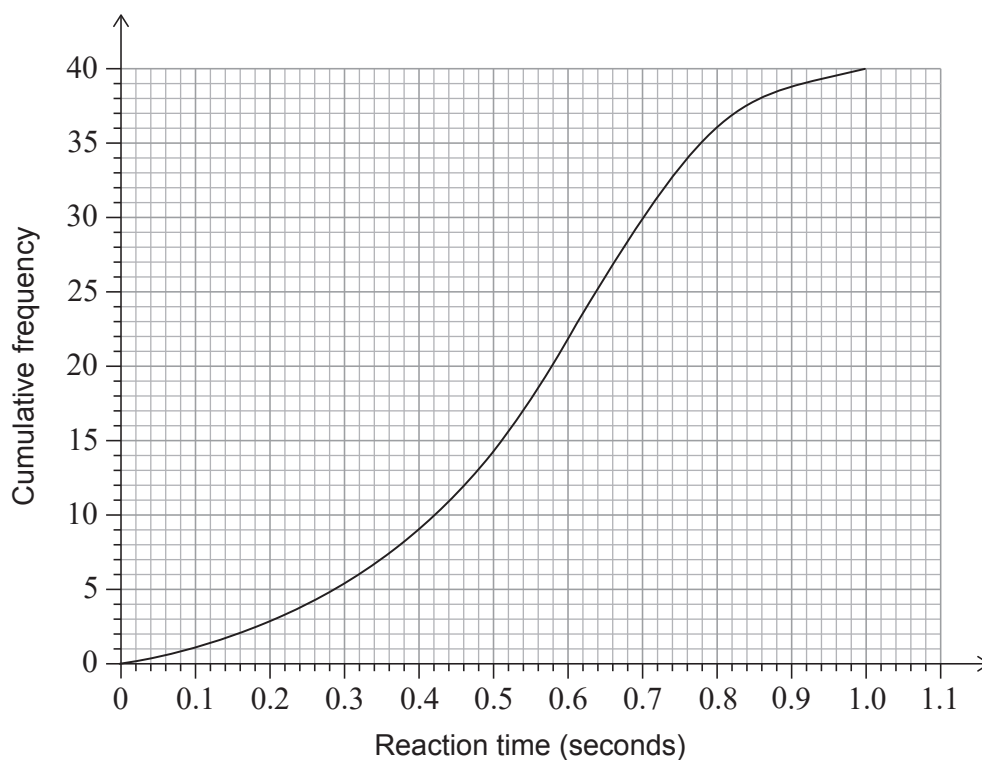
References:

2. mynamepong, n.d. Goat [image online] Available at: <https://thenounproject.com/term/goat/1761571/>
This file is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) <https://creativecommons.org/licenses/by-sa/3.0/deed.en> [Accessed 22 April 2010] Source adapted.

All other texts, graphics and illustrations © International Baccalaureate Organization 2022

1. [Maximum mark: 17]

Mackenzie conducted an experiment on the reaction times of teenagers. The results of the experiment are displayed in the following cumulative frequency graph.



- (a) Use the graph to estimate the
- (i) median reaction time;
 - (ii) interquartile range of the reaction times. [4]
- (b) Find the estimated number of teenagers who have a reaction time greater than 0.4 seconds. [2]
- (c) Determine the 90th percentile of the reaction times from the cumulative frequency graph. [2]

(This question continues on the following page)

(Question 1 continued)

Mackenzie created the cumulative frequency graph using the following grouped frequency table.

Reaction time, t (s)	Frequency
$0 < t \leq 0.2$	3
$0.2 < t \leq 0.4$	a
$0.4 < t \leq 0.6$	13
$0.6 < t \leq 0.8$	14
$0.8 < t \leq 1.0$	b

- (d) Write down the value of
- (i) a ;
 - (ii) b . [2]
- (e) Write down the modal class from the table. [1]
- (f) Use your graphic display calculator to find an estimate of the mean reaction time. [2]

Upon completion of the experiment, Mackenzie realized that some values were grouped incorrectly in the frequency table. Some reaction times recorded in the interval $0 < t \leq 0.2$ should have been recorded in the interval $0.2 < t \leq 0.4$.

- (g) Suggest how, if at all, the estimated mean and estimated median reaction times will change if the errors are corrected. Justify your response. [4]

4. [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]

[illegible]

6. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



8. [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]



Turn over

9. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



1. [Maximum mark: 17]

Elsie, a librarian, wants to investigate the length of time, T minutes, that people spent in her library on a particular day.

- (a) State whether the variable T is discrete or continuous.

[1]

Elsie's data for 160 people who visited the library on that particular day is shown in the following table.

T (minutes)	$0 \leq T < 20$	$20 \leq T < 40$	$40 \leq T < 60$	$60 \leq T < 80$	$80 \leq T < 100$
Frequency	50	62	k	14	8

- (b) Find the value of k .

[2]

- (c) (i) Write down the modal class.

- (ii) Write down the mid-interval value for this class.

[2]

- (d) Use Elsie's data to calculate an estimate of the mean time that people spent in the library. [2]

- (e) Using the table, write down the maximum possible number of people who spent 35 minutes or less in the library on that day.

[1]

Elsie assumes her data to be representative of future visitors to the library.

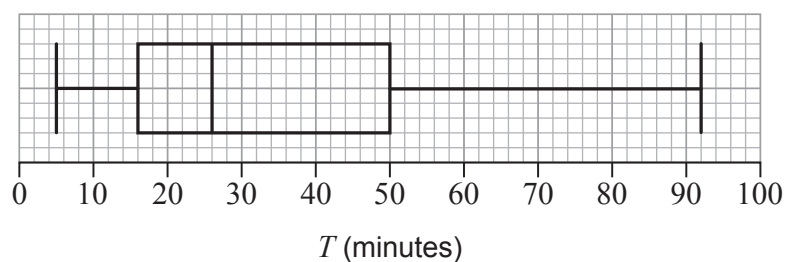
- (f) Find the probability a visitor spends at least 60 minutes in the library.

[2]

(This question continues on the following page)

(Question 1 continued)

The following box and whisker diagram shows the times, in minutes, that the 160 visitors spent in the library.



- (g) Write down the median time spent in the library. [1]
- (h) Find the interquartile range. [2]
- (i) Hence show that the longest time that a person spent in the library is not an outlier. [3]

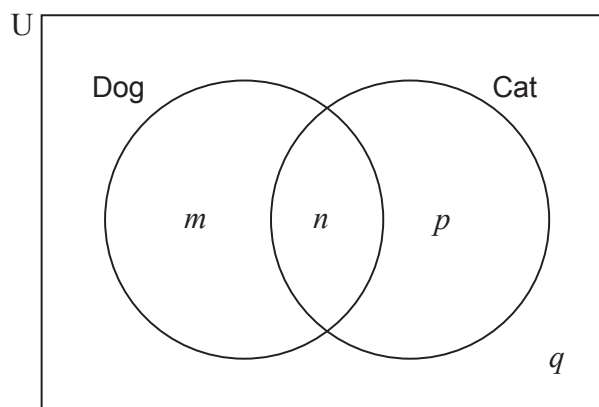
Elsie believes the box and whisker diagram indicates that the times spent in the library are not normally distributed.

- (j) Identify one feature of the box and whisker diagram which might support Elsie's belief. [1]

4. [Maximum mark: 16]

At Mirabooka Primary School, a survey found that 68% of students have a dog and 36% of students have a cat. 14% of students have both a dog and a cat.

This information can be represented in the following Venn diagram, where m , n , p and q represent the percentage of students within each region.



(a) Find the value of

(i) m .

(ii) n .

(iii) p .

(iv) q .

[4]

(b) Find the percentage of students who have a dog or a cat or both.

[1]

(c) Find the probability that a randomly chosen student

(i) has a dog but does not have a cat.

(ii) has a dog given that they do not have a cat.

[3]

(This question continues on the following page)

(Question 4 continued)

Each year, one student is chosen randomly to be the school captain of Mirabooka Primary School.

Tim is using a binomial distribution to make predictions about how many of the next 10 school captains will own a dog. He assumes that the percentages found in the survey will remain constant for future years and that the events “being a school captain” and “having a dog” are independent.

Use Tim’s model to find the probability that in the next 10 years

- (d) (i) 5 school captains have a dog.
- (ii) more than 3 school captains have a dog.
- (iii) exactly 9 school captains in succession have a dog. [7]

John randomly chooses 10 students from the survey.

- (e) State why John should not use the binomial distribution to find the probability that 5 of these students have a dog. [1]

Turn over

1. [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]

(This question continues on the following page)



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]

[illegible]

Turn over

4. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

12. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

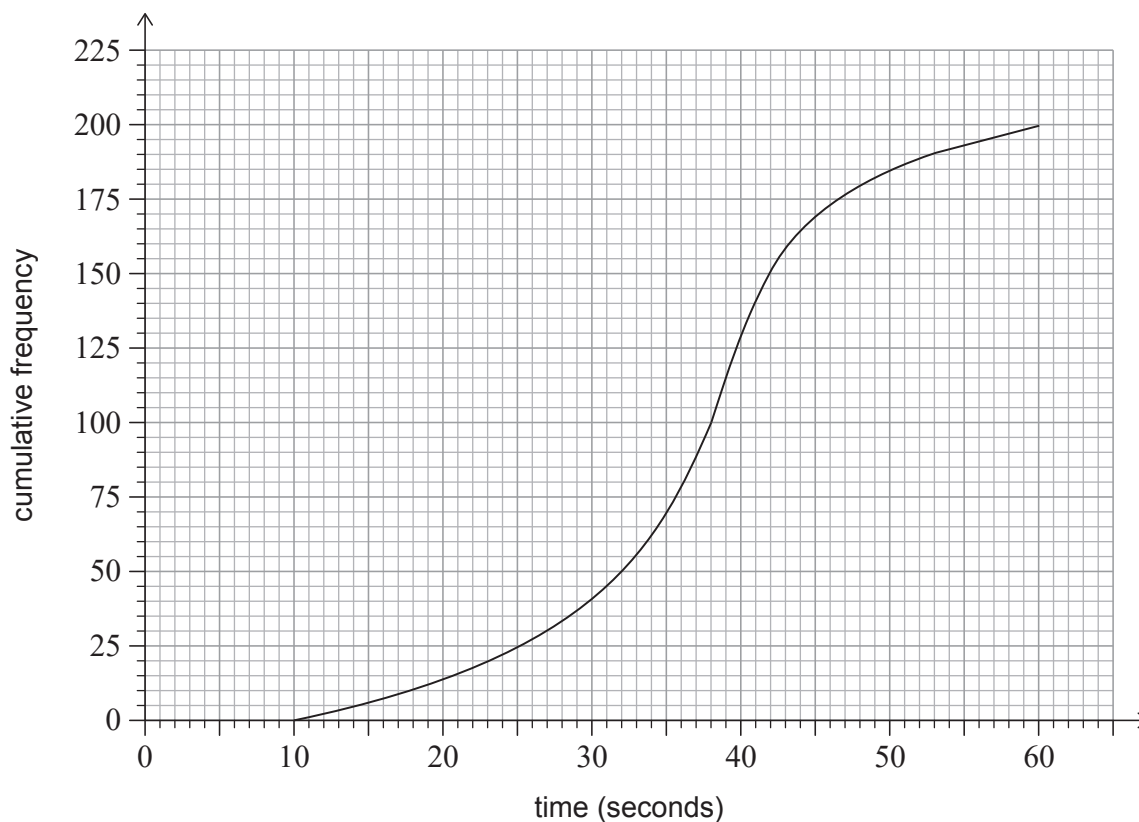
.....

.....



3. [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]

(This question continues on the following page)



(Question 3 continued)

[illegible]

Turn over

4. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



5. [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]

This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.

Turn over

6. [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]

(This question continues on the following page)



(Question 6 continued)

[illegible]

Turn over

9. [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]



Turn over

12. [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
$\mathbf{P}(X=x)$	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]

[illegible]

Turn over

1. [Maximum mark: 15]

The mean annual temperatures for Earth, recorded at fifty-year intervals, are shown in the table.

Year (x)	1708	1758	1808	1858	1908	1958	2008
Temperature $^{\circ}\text{C}$ (y)	8.73	9.22	9.10	9.12	9.13	9.45	9.76

Tami creates a linear model for this data by finding the equation of the straight line passing through the points with coordinates (1708, 8.73) and (1958, 9.45).

- (a) Calculate the gradient of the straight line that passes through these two points. [2]
- (b) (i) Interpret the meaning of the gradient in the context of the question.
- (ii) State appropriate units for the gradient. [2]
- (c) Find the equation of this line giving your answer in the form $y = mx + c$. [2]
- (d) Use Tami's model to estimate the mean annual temperature in the year 2000. [2]

Thandizo uses linear regression to obtain a model for the data.

- (e) (i) Find the equation of the regression line y on x .
- (ii) Find the value of r , the Pearson's product-moment correlation coefficient. [3]
- (f) Use Thandizo's model to estimate the mean annual temperature in the year 2000. [2]

Thandizo uses his regression line to predict the year when the mean annual temperature will first exceed 15°C .

- (g) State two reasons why Thandizo's prediction may not be valid. [2]

Turn over

4. [Maximum mark: 17]

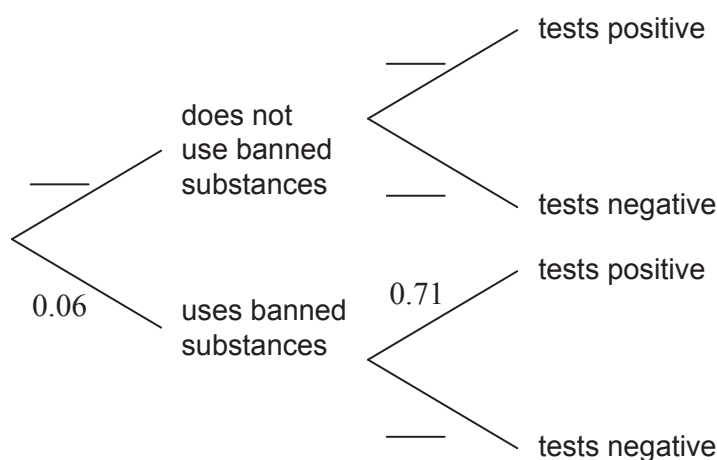
A large international sports tournament tests their athletes for banned substances. They interpret a positive test result as meaning that the athlete uses banned substances. A negative result means that they do not.

The probability that an athlete uses banned substances is estimated to be 0.06.

If an athlete **uses** banned substances, the probability that they will test positive is 0.71.

If an athlete does **not use** banned substances, the probability that they will test negative is 0.98.

- (a) Using the information given, **copy** (into your answer booklet) and complete the following tree diagram. [2]



- (b) (i) Determine the probability that a randomly selected athlete does not use banned substances and tests negative.
- (ii) If two athletes are selected at random, calculate the probability that both athletes do not use banned substances and both test negative. [4]
- (c) (i) Calculate the probability that a randomly selected athlete will receive an **incorrect** test result.
- (ii) A random sample of 1300 athletes at the tournament are selected for testing. Calculate the expected number of athletes in the sample that will receive an incorrect test result. [5]

Team X are competing in the tournament. There are 20 athletes in this team. It is known that none of the athletes in Team X use banned substances.

- (d) Calculate the probability that none of the athletes in Team X will test positive. [4]
- (e) Determine the probability that more than 2 athletes in Team X will test positive. [2]

Turn over

4. [Maximum mark: 17]

It is claimed that a new remedy cures 82 % of the patients with a particular medical problem.

This remedy is to be used by 115 patients, and it is assumed that the 82 % claim is true.

- (a) Find the probability that exactly 90 of these patients will be cured. [3]
- (b) Find the probability that at least 95 of these patients will be cured. [2]
- (c) Find the variance in the possible number of patients that will be cured. [2]

The probability that at least n patients will be cured is less than 30 %.

- (d) Find the least value of n . [3]

A clinic is interested to see if the mean recovery time of their patients who tried the new remedy is less than that of their patients who continued with an older remedy. The clinic randomly selects some of their patients and records their recovery time in days. The results are shown in the table below.

	Recovery time (days)							
Group N (new remedy)	12	13	9	13	14	15	17	

	Recovery time (days)							
Group O (old remedy)	17	11	10	18	20	22	14	15

The data is assumed to follow a normal distribution and the population variance is the same for the two groups. A t -test is used to compare the means of the two groups at the 10 % significance level.

- (e) State the appropriate null and alternative hypotheses for this t -test. [2]
- (f) Find the p -value for this test. [2]
- (g) State the conclusion for this test. Give a reason for your answer. [2]
- (h) Explain what the p -value represents. [1]

Turn over

1. [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]

(This question continues on the following page)



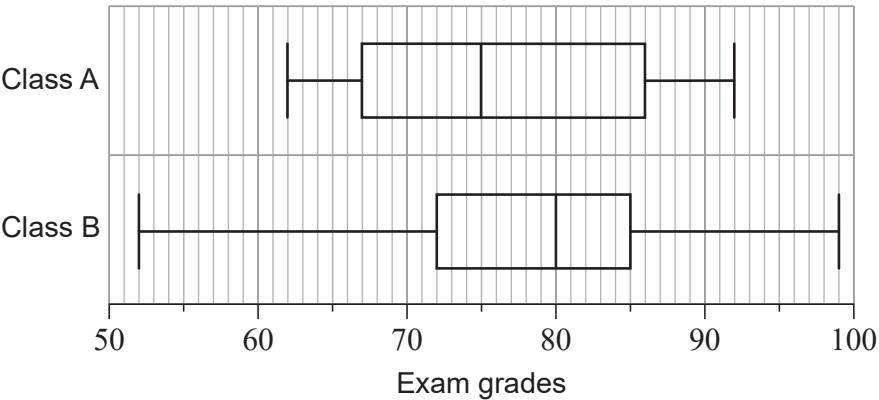
(Question 1 continued)

[illegible]

Turn over

9. [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
---------	---	---	---	----	---	---	----	---	---	---

(This question continues on the following page)



(Question 9 continued)

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]



Turn over

11. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

12. [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

(This question continues on the following page)





Disclaimer:

Content used in IB assessments is taken from authentic, third-party sources. The views expressed within them belong to their individual authors and/or publishers and do not necessarily reflect the views of the IB.

References:

2. Nina Aldin Thune. https://en.wikipedia.org/wiki/Great_Pyramid_of_Giza#/media/File:Kheops-Pyramid.jpg. Licensed under CC BY 2.5 <https://creativecommons.org/licenses/by/2.5/#>. Image adapted.

All other texts, graphics and illustrations © International Baccalaureate Organization 2023



1. [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]

(This question continues on the following page)



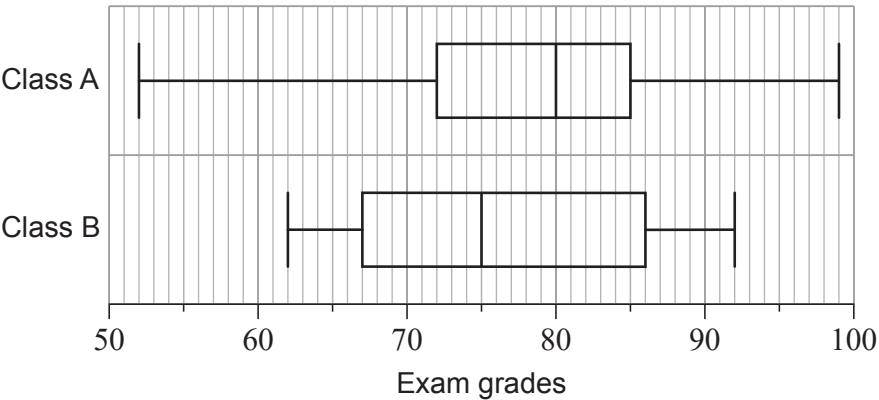
(Question 1 continued)

[illegible]

Turn over

9. [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		

(This question continues on the following page)



(Question 9 continued)

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]



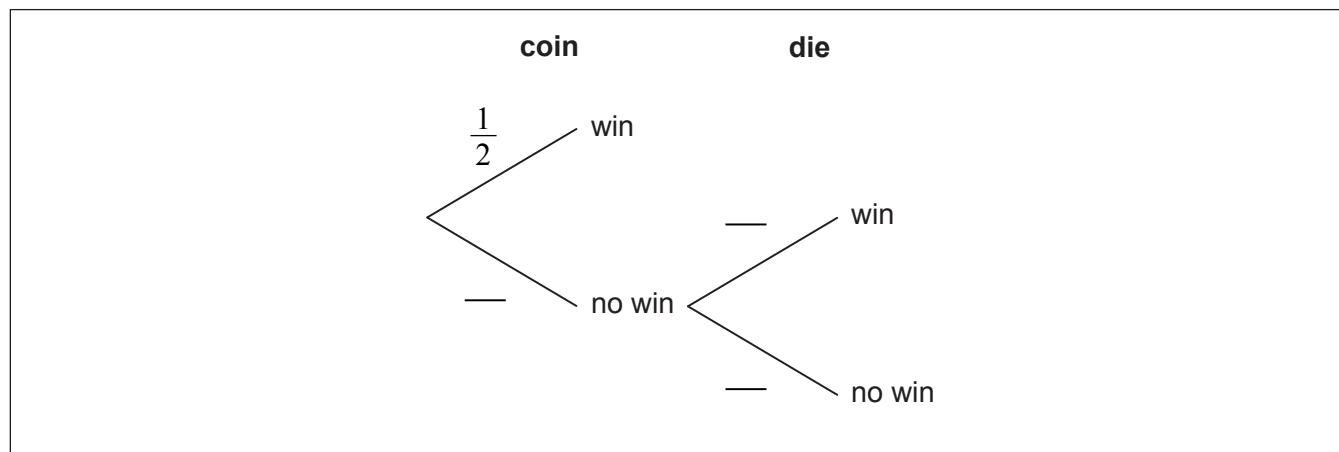
Turn over

10. [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]

[illegible]

11. [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]



Turn over

12. [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

(This question continues on the following page)



(Question 12 continued)

- (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]

This image shows a single sheet of white paper with horizontal blue or grey ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Disclaimer:

Content used in IB assessments is taken from authentic, third-party sources. The views expressed within them belong to their individual authors and/or publishers and do not necessarily reflect the views of the IB.

References:

2. Nina Aldin Thune. https://en.wikipedia.org/wiki/Great_Pyramid_of_Giza#/media/File:Kheops-Pyramid.jpg. Licensed under CC BY 2.5 <https://creativecommons.org/licenses/by/2.5/#>. Image adapted.

All other texts, graphics and illustrations © International Baccalaureate Organization 2023



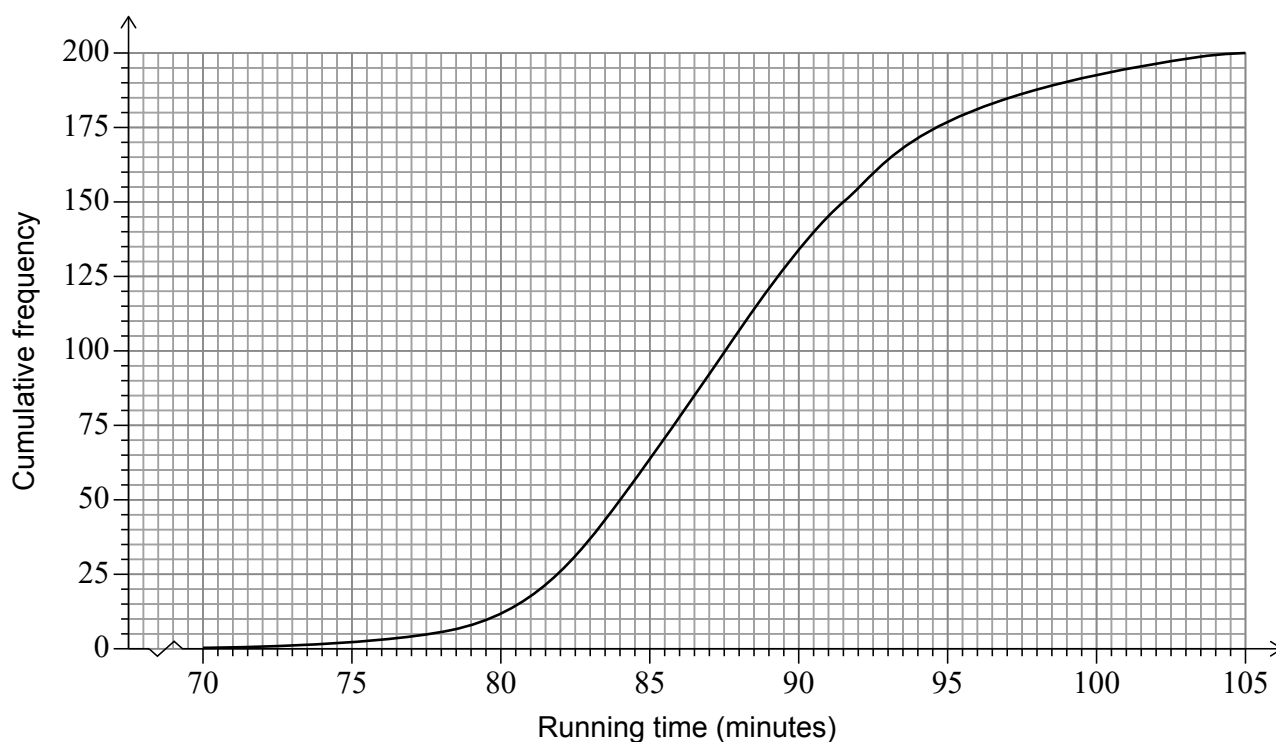
3. [Maximum mark: 18]

The running time, t (minutes), of 200 family movies are recorded in the following table.

Running time (minutes)	Frequency
$70 \leq t < 80$	11
$80 \leq t < 85$	51
$85 \leq t < 90$	68
$90 \leq t < 95$	47
$95 \leq t < 105$	23

- (a) (i) Write down the mid-interval value of $70 \leq t < 80$.
- (ii) Calculate an estimate of the mean running time of the 200 movies. [3]

This table is used to create the following cumulative frequency graph.



- (b) Use the cumulative frequency curve to estimate the interquartile range. [2]

“Star Feud” is a movie in the data set and its running time is 100 minutes.

- (c) Use your answer to part (b) to estimate whether “Star Feud’s” running time is an outlier for this data. Justify your answer. [3]

(This question continues on the following page)

(Question 3 continued)

It is believed that the running times of family movies follow a normal distribution with mean 88 minutes and standard deviation 6.75 minutes.

It is decided to perform a χ^2 goodness of fit test on the data to determine whether this sample of 200 movies could have plausibly been drawn from an underlying distribution $N(88, 6.75^2)$.

(d) Write down the null and the alternative hypotheses for the test.

[2]

As part of the test, the following table is created.

Movie running time (minutes)	Observed frequency	Expected frequency
$t < 80$	11	23.6
$80 \leq t < 85$	51	42.1
$85 \leq t < 90$	68	a
$90 \leq t < 95$	47	46.7
$95 \leq t$	23	b

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

(ii) Hence, perform the test to a 5 % significance level, clearly stating the conclusion in context.

[8]

Turn over

3. [Maximum mark: 18]

The heights, h , of 200 university students are recorded in the following table.

Height (cm)	Frequency
$140 \leq h < 160$	11
$160 \leq h < 170$	51
$170 \leq h < 180$	68
$180 \leq h < 190$	47
$190 \leq h < 210$	23

- (a)

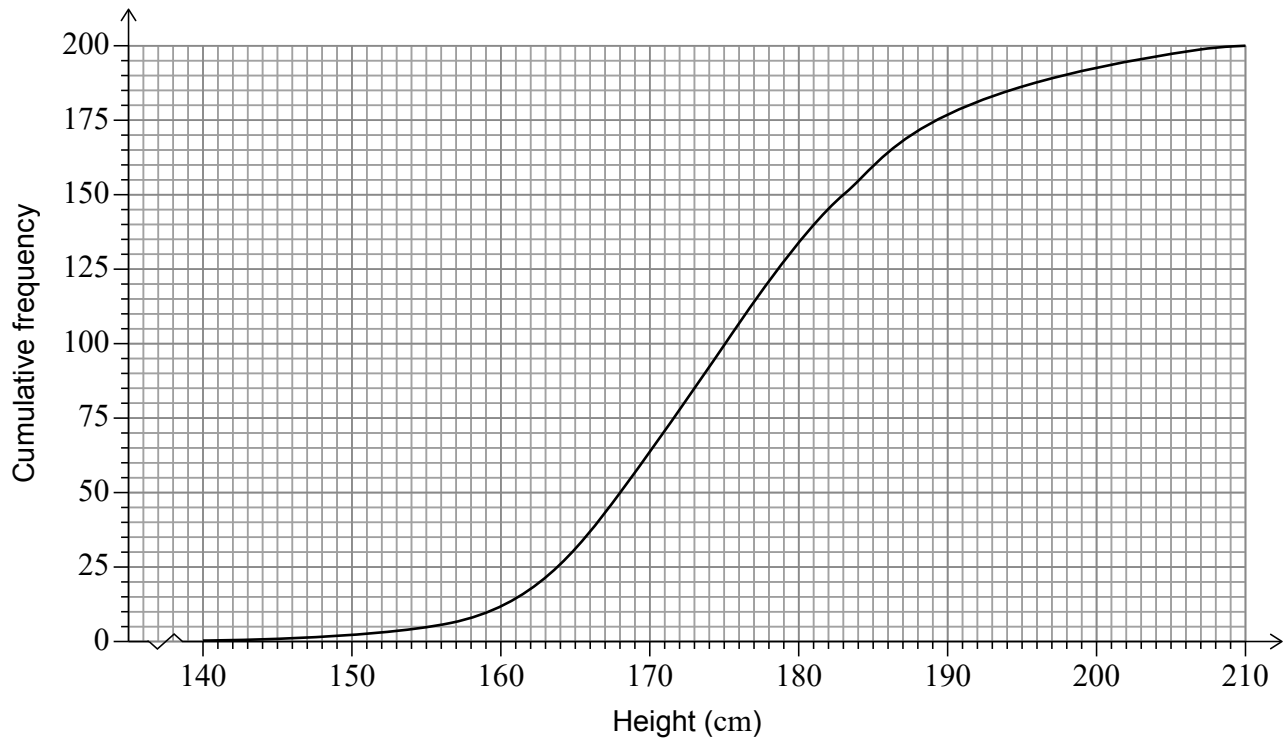
(i)

Write down the mid-interval value of $140 \leq h < 160$.
- (ii)

Calculate an estimate of the mean height of the 200 students.

[3]

This table is used to create the following cumulative frequency graph.



- (b)

Use the cumulative frequency curve to estimate the interquartile range.

[2]

Laszlo is a student in the data set and his height is 204 cm.

- (c)

Use your answer to part (b) to estimate whether Laszlo's height is an outlier for this data. Justify your answer.

[3]

(This question continues on the following page)

(Question 3 continued)

It is believed that the heights of university students follow a normal distribution with mean 176 cm and standard deviation 13.5 cm.

It is decided to perform a χ^2 goodness of fit test on the data to determine whether this sample of 200 students could have plausibly been drawn from an underlying distribution $N(176, 13.5^2)$.

(d) Write down the null and the alternative hypotheses for the test. [2]

As part of the test, the following table is created.

Height of student (cm)	Observed frequency	Expected frequency
$h < 160$	11	23.6
$160 \leq h < 170$	51	42.1
$170 \leq h < 180$	68	a
$180 \leq h < 190$	47	46.7
$190 \leq h$	23	b

(e) (i) Find the value of a and the value of b .
(ii) Hence, perform the test to a 5 % significance level, clearly stating the conclusion in context. [8]

1. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

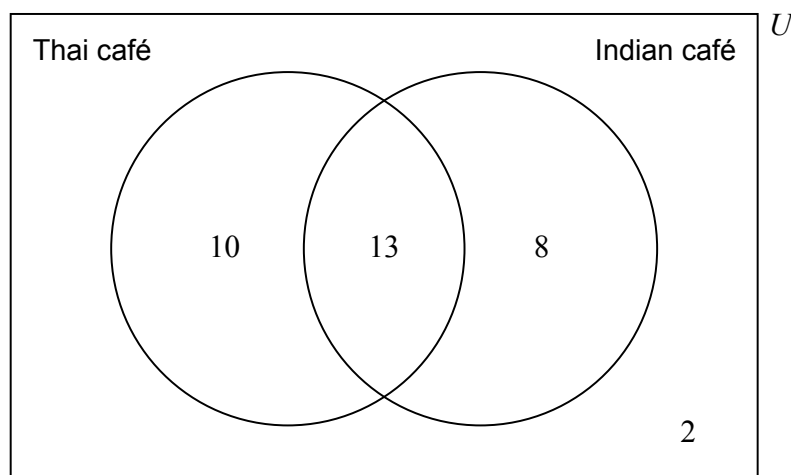
.....



Turn over

2. [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]

(This question continues on the following page)



(Question 2 continued)

[illegible]

Turn over



2227-1207

Please **do not** write on this page.

Answers written on this page
will not be marked.



8. [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]

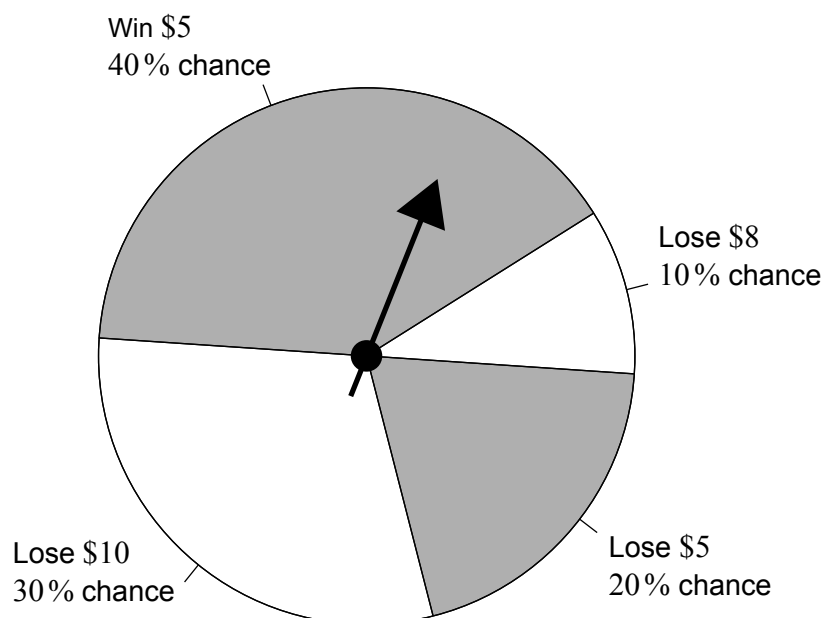


Turn over

12. [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]

(This question continues on the following page)



[illegible]

Disclaimer:

Content used in IB assessments is taken from authentic, third-party sources. The views expressed within them belong to their individual authors and/or publishers and do not necessarily reflect the views of the IB.

References:

10. Saddako, n.d. *Cheetah (Acinonyx jubatus) Running - stock photo*. [image online] Available at: <https://www.gettyimages.co.uk/detail/photo/cheetah-running-royalty-free-image/523244194?phrase=cheetah+speed&adppopup=true> [Accessed 2 May 2023]. Source adapted.
- GlobalP, n.d. *Lion, Panthera leo, 8 years old, standing - stock photo*. [image online] Available at: <https://www.gettyimages.co.uk/detail/photo/lion-panthera-leo-8-years-old-standing-royalty-free-image/134976936?phrase=Lion+standing&adppopup=true> [Accessed 2 May 2023]. Source adapted.

All other texts, graphics and illustrations © International Baccalaureate Organization 2024



1. [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]



Turn over

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

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

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

(b) (i) Calculate χ^2_{calc} .

(ii) State a conclusion to the test. Give a reason for your answer. [4]

[illegible]

Turn over

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.

- (i) two are flawed.

- [4]

[2]

[illegible]

8. [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]

[illegible]

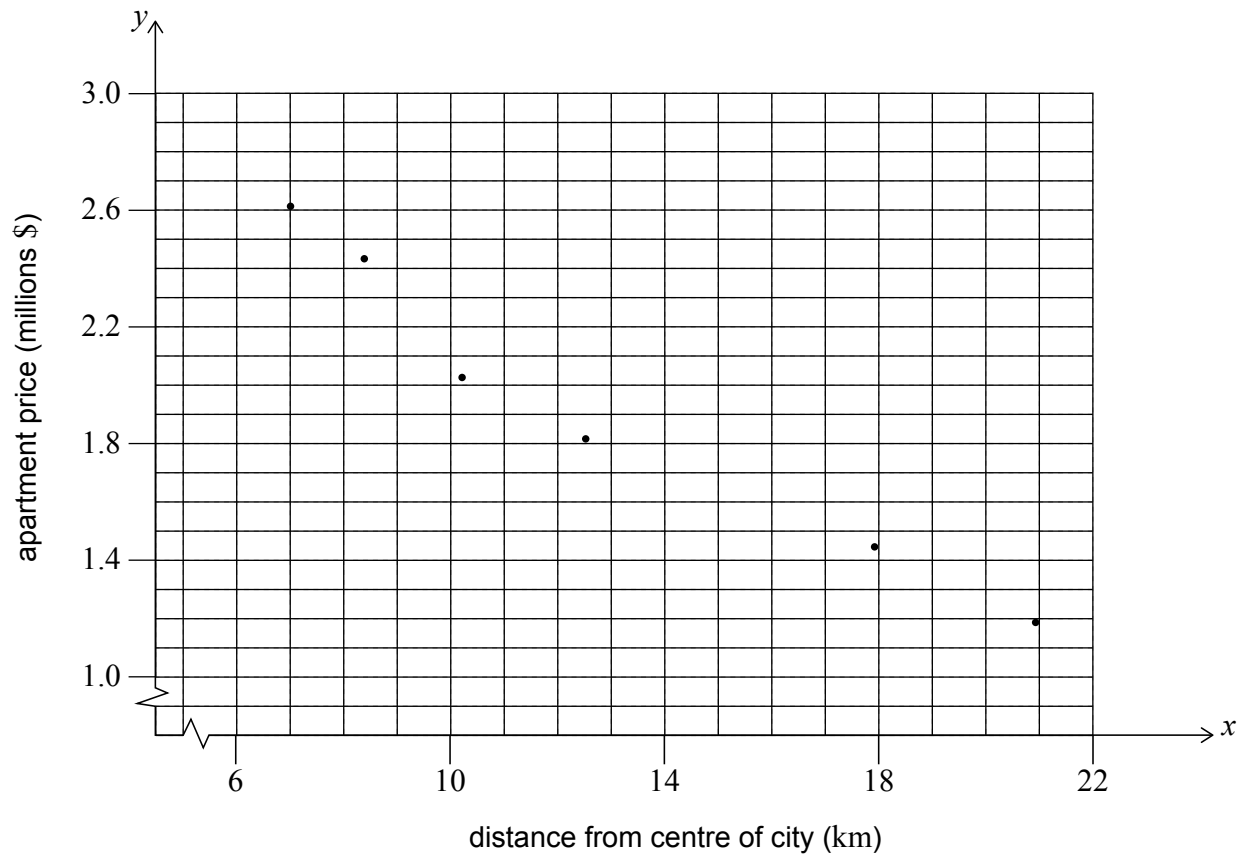
1. [Maximum mark: 19]

Xavie conducted a study to see if there is a relationship between the price of an apartment, y , and its distance, x , from the city centre of Melbourne.

They took a random sample of six typical apartments along a train line in the city. Xavie obtained the data shown in the following table.

x (kilometres)	7.0	8.4	10.3	12.5	17.8	20.9
y (millions of dollars)	2.61	2.44	2.03	1.81	1.45	1.18

A plot of these data is seen in the following graph.



(This question continues on the following page)

(Question 1 continued)

- (a) Write down the value of the Spearman’s rank correlation coefficient, r_s . [1]
- (b) (i) Find the Pearson’s product-moment correlation coefficient, r .
- (ii) Use your value of r to state which **two** of the following would best describe the correlation between the variables.

Positive	Negative	Strong	Weak	No correlation	[4]
----------	----------	--------	------	----------------	-----

The relationship between the variables can be modelled by the regression equation $y = ax + b$.

- (c) (i) Write down the value of a .
- (ii) Write down the value of b .
- (iii) According to this model, state in context what the value of b represents. [3]
- (d) Xavie uses the regression equation to estimate the price of a typical apartment located 19.6 km from the city centre.
- (i) Find this estimated price.
- (ii) State two reasons that Xavie might use to justify the validity of this estimate. [5]

(This question continues on page 5)

— T —

2227-1200

Blank page

(Question 1 continued)

To verify whether this relationship applies in a different direction from the city centre, Xavie considers two locations, A and B, both an equal distance from the city centre. They take a random sample of seven apartments from each location and record the prices (in millions of dollars) in the following tables.

Apartment price in location A						
1.21	1.25	1.31	1.32	1.58	1.95	2.13

Apartment price in location B						
1.51	1.58	1.69	2.61	2.72	2.81	2.95

Xavie conducts a t -test, at the 5 % level of significance, to see if the mean apartment price in location A is different to the mean apartment price in location B. They assume the population variances are the same.

For this test, Xavie takes the null hypothesis to be $\mu_A = \mu_B$.

- (e) Write down the alternative hypothesis. [1]
- (f) Find the p -value for this test. [2]
- (g) State the conclusion of the test. Justify your answer. [2]
- (h) State one additional assumption Xavie has made about the distributions to conduct this test. [1]

2. [Maximum mark: 15]

The company Fred Express delivers packages. From past experience, the time taken, T , to deliver a package follows a normal distribution with mean 64 hours and standard deviation 12 hours.

(a) State $P(T < 64)$. [1]

(b) Find $P(44 < T < 64)$. [2]

30% of packages are delivered in less than k hours.

(c) (i) Sketch a diagram of this normal distribution, shading the region that represents $P(T < k)$.

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

For quality control, the manager randomly selects five outgoing packages. These selections are independent.

(d) Find the probability that exactly two of these packages are delivered in less than k hours. [3]

Fred Express charges a fixed amount of \$4.50 for any package weighing 1 kg or less. Heavier packages are charged an additional fee of \$2.00 per kg. This fee is applied for any weight **in excess of** 1 kg. For example, a 1.5 kg package is charged an additional \$1.00.

(e) Write down an expression for the amount charged to deliver a package of weight x kg, where $x > 1$. [2]

(f) Find the amount Fred Express charges for a 5.3 kg package. [1]

Meiling is charged \$7.20 for the delivery of a package.

(g) Find the weight of Meiling's package. [2]

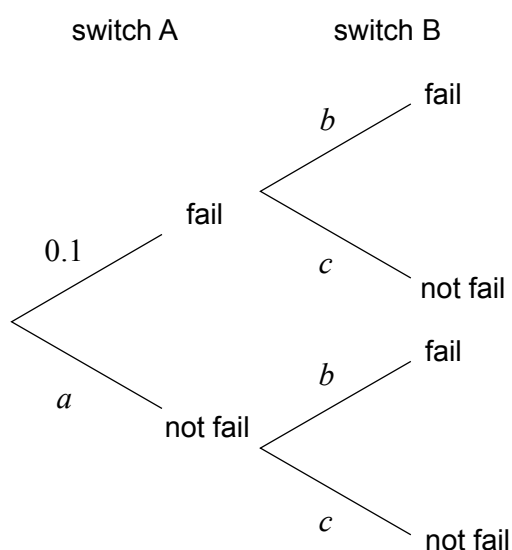
4. [Maximum mark: 12]

A type of generator will only function if a particular switch is working. The generator has a main switch, A, and a 'back up' switch, B.

The manufacturer claims the probability of switch A failing within one month of being fitted is 0.1 and the probability of the cheaper switch B failing within one month is 0.3. Whether or not a switch fails is independent of the state of the other switch.

If both switches fail, the generator needs to shut down to replace the switches. Both switches are replaced after a month of use (whether they have failed or not) or whenever the generator needs to be shut down.

The following tree diagram shows the probabilities of a switch failing within one month of them both being replaced, assuming the manufacturer's claim is correct.



(a) Write down the values of

(i) a .

(ii) b .

(iii) c .

[2]

(b) Hence find the probability that the generator needs to shut down within one month of the switches being replaced.

[1]

(This question continues on the following page)

(Question 4 continued)

The owner of the generator is suspicious of the switch manufacturer’s claims, so they look back through the past 200 occasions when the switches were replaced. The records show whether no switches, one switch or two switches had failed.

The data the owner collected are shown in the following table.

No switch fails	One switch fails	Two switches fail
118	72	10

- (c)

Show that the expected value of no switches failing in the generator, during the last 200 occasions when the switches were replaced, is 126.

[2]
- (d)

Perform a χ^2 goodness of fit test at the 5% significance level to test whether the manufacturer’s claims are correct using the following hypotheses.

H_0 : The manufacturer’s claims are correct.

H_1 : The manufacturer’s claims are not both correct.

[7]

1. [Maximum mark: 16]

In a given week, the number of students in a particular primary school that were absent due to headlice (H), influenza (I) and/or chickenpox (C) were recorded as follows.

The primary school has 500 students.

35	students had headlice only
20	students had influenza only
5	students had chickenpox only
4	students had headlice and influenza but not chickenpox
2	students had headlice and chickenpox but not influenza
3	students had influenza and chickenpox but not headlice
1	student had headlice, influenza and chickenpox

(a) Draw a Venn diagram to represent this information. [3]

(b) Calculate the number of students who did not have headlice or influenza or chickenpox. [2]

A student is chosen at random from all the students in the school.

(c) Find the probability that this student has

(i) headlice.

(ii) influenza given that the student has headlice. [4]

(This question continues on the following page)

(Question 1 continued)

Diego is a teacher in the school. He believes that the number of students, n , who have had influenza during the first t days of the school year, can be modelled by the function

$$n(t) = 250 - 240(2)^{kt}, \quad k \in \mathbb{R}.$$

- (d) Use Diego's model to calculate the number of students who started the school year with influenza. [2]

It is known that 130 students have had influenza during the first 10 days of the school year.

- (e) Find the value of k . [2]
- (f) Using this model, calculate how many days it will take for 200 students to have had influenza since the start of the school year. [2]

By the last day of the school year, it is known that 300 students have had influenza.

- (g) Comment on the appropriateness of Diego's model. [1]

Turn over

4. [Maximum mark: 19]

A recent study found that the heights of Dutch women can be modelled by a normal distribution with mean 170.7 cm and standard deviation 6.3 cm.

A Dutch woman is chosen at random.

(a) Calculate the probability that her height is

(i) less than 160 cm.

(ii) between 160 cm and 170 cm.

[4]

27% of Dutch women have a height of more than h metres.

(b) Calculate the value of h .

[2]

Janneke selects a random sample of 200 Dutch women from Amsterdam and measures their heights. She wants to determine whether this sample could have been chosen from a normally distributed population with mean of 170.7 cm and standard deviation of 6.3 cm.

She performs a χ^2 goodness of fit test at the 5% significance level. She begins by creating the following frequency table.

Height, h	Observed frequency	Expected frequency
$h \leq 160$	6	8.943
$160 < h \leq 170$	85	a
$170 < h \leq 180$	92	b
$h > 180$	17	13.99

(c) Calculate, correct to four significant figures, the value of

(i) a .

(ii) b .

[2]

(This question continues on the following page)

(Question 4 continued)

The hypotheses for Janneke’s test are

H_0 : the heights are drawn from a normally distributed population with mean 170.7 cm and standard deviation 6.3 cm

H_1 : the heights are not drawn from a normally distributed population with mean 170.7 cm and standard deviation 6.3 cm

(d) Write down the degrees of freedom for this test. [1]

The critical value for this test is 7.815.

(e) Perform the χ^2 goodness of fit test and state your conclusion, justifying your reasoning. [4]

Gundega claims that, on average, Latvian women are taller than Dutch women.

Random samples of 10 Latvian women and 10 Dutch women are chosen, and their heights are measured.

Heights of Latvian women (cm)									
171	163	180	159	169	182	166	168	171	170

Heights of Dutch women (cm)									
173	182	181	166	175	161	169	165	172	169

Gundega performs a t -test at the 5% significance level. It is assumed that the populations are normally distributed and have equal variances.

(f) Write down the null and alternative hypotheses for this test. [2]

(g) Perform the t -test and state the conclusion, justifying your reasoning. [4]

2. [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

- Find the mean **and** standard deviation of these test scores. Give your answers correct to four significant figures. [3]
- The scores are multiplied by ten. Write down the new mean and standard deviation. [2]
- 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]

[illegible]

Turn over

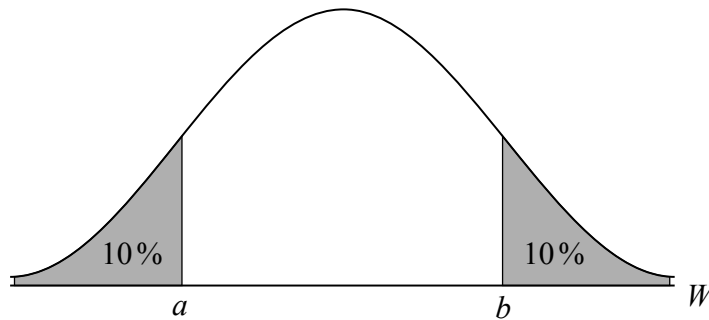
7. [Maximum mark: 7]

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

- (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]

[illegible]

9. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



11. [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]



2. [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]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

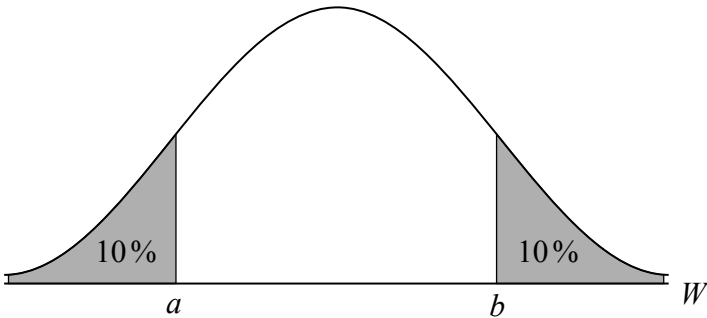
7. [Maximum mark: 7]

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 .
- (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.

[3]

[2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



9. [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]



11. [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]



3. [Maximum mark: 14]

A survey was answered by 20 000 expatriates (people living in a country that is not their own). The data ranked countries in order of the country they felt was best for expatriates. The highest-ranked country was Switzerland.

These results were compared to happiness scores taken from *The World Happiness Report 2022*. The following table shows this data for the top 10 expatriate countries.

Country	Switzerland	New Zealand	Spain	Australia	Cyprus	Portugal	Ireland	United Arab Emirates	France	Netherlands
Expatriate country rank	1	2	3	4	5	6	7	8	9	10
Happiness score	7.5	7.2	6.5	7.2	6.2	6.0	7.0	6.6	6.7	7.4

- (a) For the **happiness score**, find
- (i) the upper quartile

(ii) the interquartile range.
- [4]
- (b) Show that Switzerland’s happiness score is not an outlier for this data.
- [3]

(This question continues on the following page)

(Question 3 continued)

The happiness scores were ranked to calculate Spearman’s rank correlation coefficient, r_s . These ranks are shown in the following table.

Country	Switzerland	New Zealand	Spain	Australia	Cyprus	Portugal	Ireland	United Arab Emirates	France	Netherlands
Happiness score	7.5	7.2	6.5	7.2	6.2	6.0	7.0	6.6	6.7	7.4
Happiness rank	1	a	b	c	9	10	5	7	6	2

- (c)

Write down the value of

(i)

a

(ii)

b

(iii)

c .
- [3]
- (d)

(i)

Find r_s .

(ii)

If France's happiness score is upgraded to 6.9, explain why the value of r_s does not change.
- [3]
- Jose concludes from this data that countries with high happiness scores are likely to be favourite expatriate countries.

(e)

State, with a reason, whether Jose’s conclusion is appropriate.
- [1]

3. [Maximum mark: 14]

A survey was answered by 20 000 expatriates (people living in a country that is not their own). The data ranked countries in order of the country they felt was best for expatriates. The highest-ranked country was Switzerland.

These results were compared to happiness scores taken from *The World Happiness Report 2022*. The following table shows this data for the top 10 expatriate countries.

Country	Switzerland	New Zealand	Spain	Australia	Cyprus	Portugal	Ireland	United Arab Emirates	France	Netherlands
Expatriate country rank	1	2	3	4	5	6	7	8	9	10
Happiness score	7.5	7.2	6.5	7.2	6.2	6.0	7.0	6.6	6.7	7.4

- (a) For the **happiness score**, find
- (i) the upper quartile

(ii) the interquartile range.
- [4]
- (b) Show that Switzerland’s happiness score is not an outlier for this data.
- [3]

(This question continues on the following page)

(Question 3 continued)

The happiness scores were ranked to calculate Spearman’s rank correlation coefficient, r_s . These ranks are shown in the following table.

Country	Switzerland	New Zealand	Spain	Australia	Cyprus	Portugal	Ireland	United Arab Emirates	France	Netherlands
Happiness score	7.5	7.2	6.5	7.2	6.2	6.0	7.0	6.6	6.7	7.4
Happiness rank	1	a	b	c	9	10	5	7	6	2

- (c)

Write down the value of

(i)

a

(ii)

b

(iii)

c .
- [3]
- (d)

(i)

Find r_s .

(ii)

If France's happiness score is upgraded to 6.9, explain why the value of r_s does not change.
- [3]

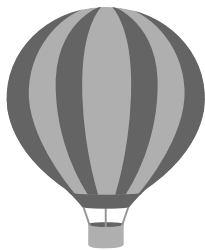
Jose concludes from this data that countries with high happiness scores are likely to be favourite expatriate countries.

- (e)

State, with a reason, whether Jose’s conclusion is appropriate.
- [1]

1. [Maximum mark: 6]

Hot air balloons are available in many sizes for different numbers of passengers.



The following table shows the recommended minimum volume of a hot air balloon, in cubic metres, for a specific number of passengers.

Number of Passengers (x)	1	2	4	5	7	15
Recommended Minimum Volume (y)	1000	1350	2225	2500	3150	5800

- (a) Use your graphic display calculator to find the Pearson’s product-moment correlation coefficient, r , for these values.

[2]
- (b) (i) Find the equation of the regression line y on x for this data in the form $y = ax + b$.

(ii) State what the value of a means in the context of the question.

[2]
- (c) Use your regression equation from part (b)(i) to find the recommended minimum volume of a balloon for 10 passengers.

[2]

(This question continues on the following page)



(Question 1 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

2. [Maximum mark: 5]

A car manufacturer collected data about preferred car colour from drivers in various age groups. The results are presented in the following table.

	Preferred car colour				Total
	White	Black	Silver	Red	
$18 \leq \text{Age} < 25$	12	7	4	17	40
$25 \leq \text{Age} < 45$	15	b	10	12	58
$\text{Age} \geq 45$	12	18	16	6	52
Total	39	46	30	35	150

(a) Write down the value of b .

[1]

The car manufacturer performs a χ^2 test for independence, at the 1% significance level, to determine if there is significant evidence that different age groups have different car colour preferences.

The null hypothesis and alternative hypothesis are defined as:

H_0 : age and car colour preference are independent.

H_1 : age and car colour preference are not independent.

The χ^2 critical value for this test is 16.81.

(b) Find the χ^2 test statistic.

[2]

(c) Write down the conclusion to the test, in context. Give a reason for your answer.

[2]

(This question continues on the following page)



(Question 2 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

7. [Maximum mark: 5]

Children inherit genes from their biological parents that determine their eye colour.

Maria and Alex plan to have 5 children. The probability that a child born to Maria and Alex will have brown eyes is 0.75. It is assumed that a child's eye colour is independent of any other child's eye colour.

Maria and Alex wish to predict how many of their 5 children will have brown eyes.

- (a) State one criterion, in addition to independence, that would support them using a binomial distribution to find probabilities for their prediction.

[1]

- (b) Calculate the probability that

- (i) exactly 3 of the 5 children will have brown eyes.

- (ii) at least 4 of the 5 children will have brown eyes.

[4]

[illegible]

2. [Maximum mark: 8]

Two unbiased dice, each with faces numbered from 1 to 6 inclusive, are rolled. The numbers on the uppermost faces of the dice are noted.

Let the random variable X be the difference between the numbers on the dice.

- (a)

(i)

Find $P(X = 0)$.
- (ii)

Find $P(X = 5)$.

[2]
- (b)

Complete the table to show the probability distribution of X .

[2]

The probability that $X = 3$ is shown.

x	0	1	2	3	4	5
$P(X = x)$				$\frac{6}{36}$		

- (c)

Calculate $E(X)$.

[2]
- (d)

Given that the difference between the numbers on the dice is an odd number, find the probability that $X = 5$.

[2]

(This question continues on the following page)



(Question 2 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over



222071314

Please **do not** write on this page.

Answers written on this page
will not be marked.



4. [Maximum mark: 8]

Jane tested three different flavours of pet food on cats and dogs to determine whether the preferred flavour was independent of the type of animal.

The results are given in the table.

	Preferred flavour		
	Beef	Chicken	Fish
Cat	8	22	20
Dog	16	25	9

Perform a χ^2 test at both the 5% and 1% levels to investigate the results.

- (a) State the null and alternative hypotheses. [1]
- (b) Write down the number of degrees of freedom. [1]
- (c) Calculate the p -value for the test. [2]
- (d) State the conclusion to the test, giving your reason
- (i) at the 1% level.

(ii) at the 5% level. [3]

Jane’s conclusion to her test was that the preferred flavour was not independent of the type of animal.

- (e) State which of the two significance levels Jane used in her test. [1]

(This question continues on the following page)



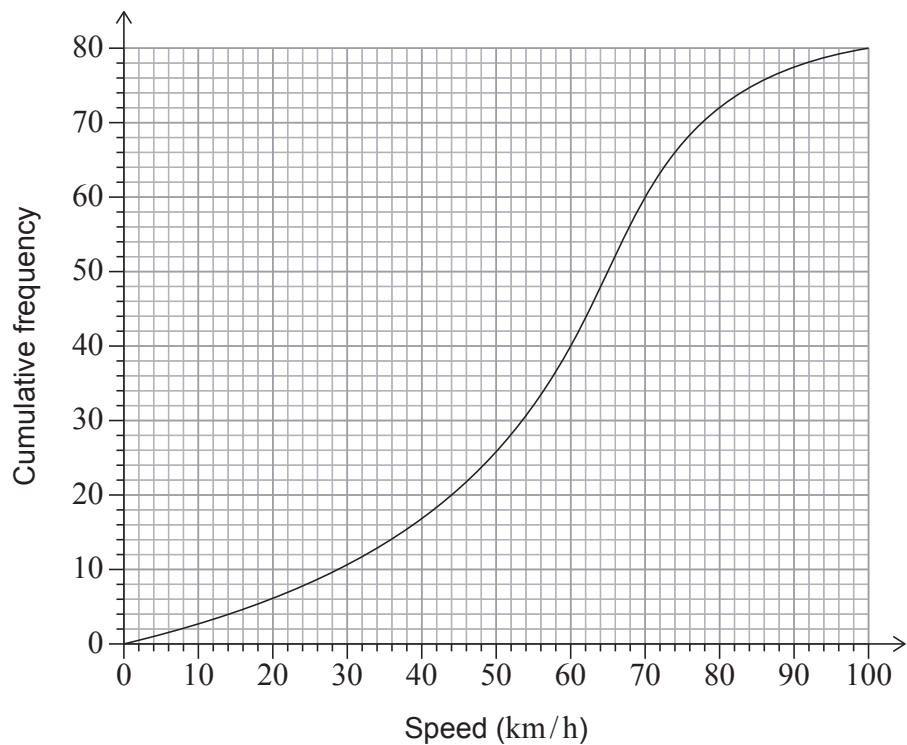
(Question 4 continued)

[illegible]

Turn over

8. [Maximum mark: 8]

The cumulative frequency curve shows the speeds in kilometres per hour of 80 cars along a road.



(a) Calculate the percentage of these cars that exceed the speed limit of 80km/h. [2]

Consider the grouped frequency table constructed using the cumulative frequency curve.

Speed (x)(km/h)	$0 < x \leq p$	$p < x \leq m$	$m < x \leq q$	$q < x \leq 100$
Number of cars	20	20	20	20

- (b) Find the value of
- (i) p .
 - (ii) m .
 - (iii) q . [3]
- (c) Hence calculate an estimate of the mean speed of these cars. [3]

(This question continues on the following page)



(Question 8 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

10. [Maximum mark: 6]

Two judges, Brett and Clarence, rank the skill levels of eight sheepdogs in a competition. The sheepdogs are labelled A to H and the judges rank the dogs as shown in the table.

Rank	1	2	3	4	5	6	7	8
Brett	A	C	D	B	E	F	G	H
Clarence	A	B	D	C	E	G	F	H

- (a) Write down the rank that Brett awards sheepdog B. [1]
- (b) Calculate Spearman's rank correlation coefficient for these data. [4]
- (c) Comment on your answer to part (b) in terms of the ranks awarded by Brett and Clarence. [1]

[illegible]

Turn over

2. [Maximum mark: 5]

The monthly energy consumption, x , in kilowatt hours (kWh) of 150 households in Helvetia is shown in the following table.

Energy consumption (x kWh)	Number of households
$500 \leq x < 700$	38
$700 \leq x < 900$	45
$900 \leq x < 1100$	25
$1100 \leq x < 1300$	18
$1300 \leq x < 1500$	16
$1500 \leq x < 1700$	8

- (a) For this data set, use the mid-point interval values to find estimates for
- (i) the mean.
 - (ii) the standard deviation.

[4]

The standard deviation of the monthly energy consumption in another nearby residential area, Eureka, is found to be 95 kWh.

- (b) Interpret the meaning of the value of the standard deviation in Eureka in comparison with the standard deviation in Helvetia.

[1]



8. [Maximum mark: 7]

Prakash is the leader of a customer service team and is interested in determining whether there is a relationship between a customer's satisfaction level and the type of service interaction they have experienced.

He collects data from a random sample of 250 customers and tracks their satisfaction level after three types of service interactions: in-person, online chat bots and website contact forms.

He categorizes the satisfaction levels as satisfied, neutral and dissatisfied.

He records the data in the following table.

		Satisfaction level		
		Satisfied	Neutral	Dissatisfied
Type of service interaction	In-person	35	30	23
	Online chat bots	31	39	23
	Website contact forms	19	28	22

Prakash performs a χ^2 test for independence at the 5% significance level.

The critical value is 9.488.

The null hypothesis, H_0 , is the satisfaction level and the type of service interaction are independent.

(a) State the alternative hypothesis for this test. [1]

(b) Find the degrees of freedom for this test. [1]

(c) Find χ^2_{calc} , the chi-squared test statistic. [2]

Prakash concludes that there is sufficient evidence to reject the null hypothesis.

(d) (i) State whether Prakash is correct. Justify your answer.

(ii) Write down the conclusion for this test in context. [3]

(This question continues on the following page)



(Question 8 continued)

[illegible]

Turn over

Please **do not** write on this page.

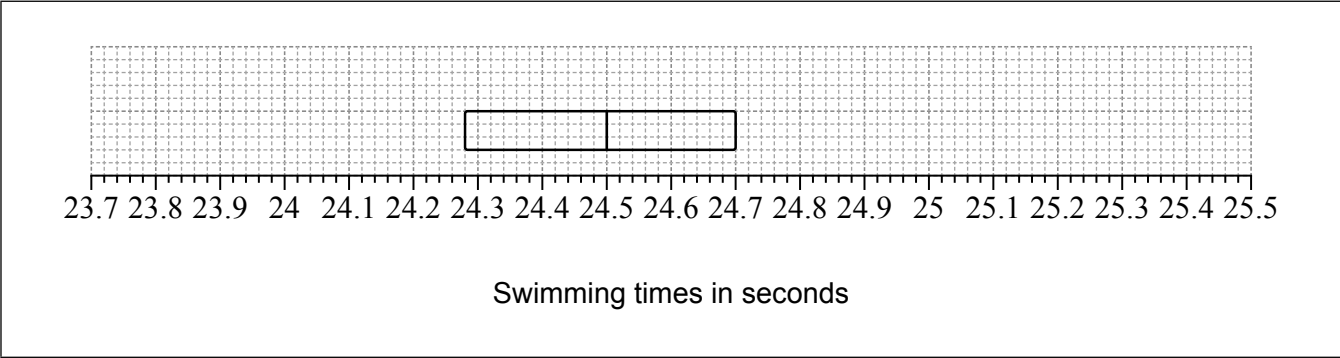
Answers written on this page
will not be marked.



10. [Maximum mark: 7]

The times, in seconds, for the fastest 16 women in a 50 m freestyle swimming championship event were recorded. All swimmers recorded different times.

Part of a box and whisker diagram for these times is shown in the following diagram.



- (a) Write down the number of swimmers who took more than 24.70 seconds to complete the race. [1]
- (b) Find the interquartile range (IQR) for the data. [2]

An outlier is defined as a value that satisfies one of the following:

- more than $1.5 \times \text{IQR}$ below the lower quartile
- more than $1.5 \times \text{IQR}$ above the upper quartile.

Of the 16 women, the two fastest swimmers took 23.96 and 24.12 seconds and the two slowest women took 25.12 and 25.40 seconds to complete the race.

- (c) (i) Show that only one of these times is an outlier.
- (ii) Complete the box and whisker diagram above. [4]

(This question continues on the following page)



(Question 10 continued)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Turn over

Please **do not** write on this page.

Answers written on this page
will not be marked.



12. [Maximum mark: 7]

A team of researchers is using a model to predict the relative happiness of different countries. To do this, a value x is calculated based on easily measured parameters, for example, life expectancy, or available social support. It is assumed that higher values of x indicate greater happiness.

To test the model a survey is conducted in six countries, A, B, C, D, E and F. In these countries the level of happiness is assessed directly using questionnaires and given a score y , out of 10, with higher scores indicating greater happiness.

To select the countries for the survey, all countries are divided into three equal groups based on wealth and two countries are chosen randomly from each group.

(a) Write down the name of this type of sampling. [1]

The results of the survey, along with the value obtained from the model, are given in the following table.

Country	A	B	C	D	E	F
Value from the model (x)	12.3	15.2	14.1	18.5	20.1	19.2
Happiness score (y)	5.2	7.3	6.2	6.9	8.0	7.2

The researchers will accept the model is a valid predictor of happiness score if the Pearson's product-moment correlation coefficient, r , is greater than 0.8.

(b) (i) Find the value of r .
(ii) Hence state whether the model can be regarded as a valid predictor of happiness score. [3]

(c) Find the equation of the regression line y on x . [1]

For a particular country $x = 17.2$.

(d) Use the regression line to predict the happiness score for this country. [2]

(This question continues on the following page)



(Question 12 continued)

This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.

Please **do not** write on this page.

Answers written on this page
will not be marked.



Please **do not** write on this page.

Answers written on this page
will not be marked.



Please **do not** write on this page.

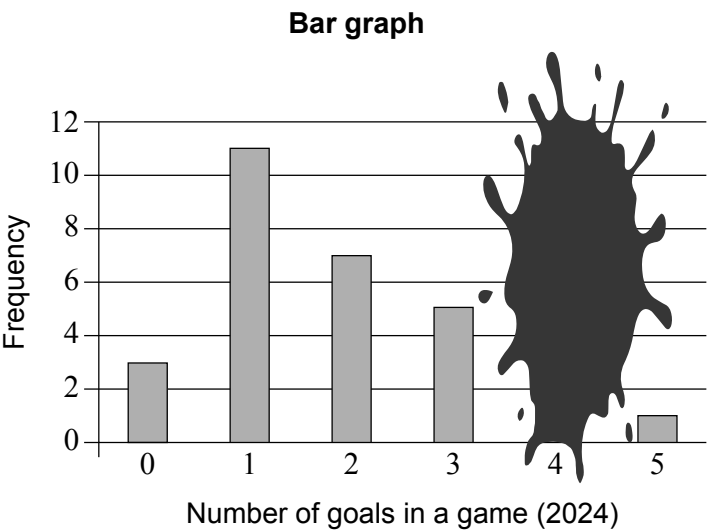
Answers written on this page
will not be marked.



1. [Maximum mark: 15]

Paul has a bar graph for the total number of goals scored in each game of a soccer tournament in 2024. The bar graph is shown below, however the frequency of 4 goals in a game is unreadable.

Paul uses this bar graph to create a frequency table.



Frequency table

Number of goals in a game (2024)	Frequency
0	3
1	11
2	7
3	k
4	p
5	1

(a) Write down the value of k . [1]

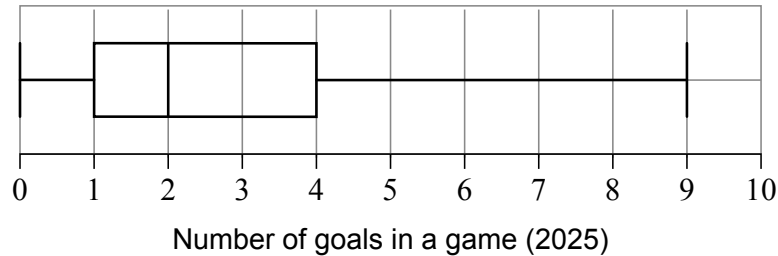
Paul knows that the mean number of goals per game scored during the tournament was 2.2.

(b) (i) Write down an equation for the mean in terms of p .
(ii) Determine the value of p . [3]

(This question continues on the following page)

(Question 1 continued)

Data for the number of goals per game in the **2025 soccer tournament** are shown in the following box and whisker diagram.



After comparing the box and whisker diagram from the 2025 tournament with the frequency table from the 2024 tournament, Paul concludes that the distribution of goals is consistent between the two tournaments.

- (c) State two observations that support Paul’s conclusion using values from the data to compare any **two** of:

range, symmetry, median, and interquartile range. [3]

Paul plans to watch all the games from the **2024** tournament in a random order. He will watch each game once.

For the first game he watches, he defines event F as:

“scoring either 0 goals or exactly 1 goal”.

- (d) Write down the event(s) from the table that are equivalent to F' . There may be more than one correct event. [2]

Event	Description
A	Scoring exactly 2 goals in this game
B	Scoring more than 1 goal in this game
C	Scoring at least 2 goals in this game
D	Scoring either 0 goals or exactly 1 goal in all games except this game
E	Not scoring 0 goals or exactly 1 goal in any game

- (e) If exactly 1 goal was scored in the first game Paul watches, write down the probability that exactly 1 goal was scored in the second game he watches. Give your answer as a fraction. [2]

- (f) Calculate the probability that 5 goals were scored in the first game that Paul watches **and** 0 goals were scored in the second game he watches. [4]

Turn over

3. [Maximum mark: 20]

According to the timetable, the evening train from Cardiff arrives in London at 6:00 pm. Records show that the arrival times for this train are normally distributed with a mean arrival time of 6:20 pm and a standard deviation of 10 minutes.

Let X be the random variable “the number of minutes late the train arrives”.

(a) Write down

(i) the mean of X .

(ii) the variance of X .

[2]

(b) Find the probability that the evening train arrives in London between 15 and 25 minutes late.

[2]

If the train arrives more than 30 minutes late, the train company must pay a fine.

(c) Calculate the expected number of days, in a 365-day year, that the train company pays a fine.

[3]

Janine arrives in London on the evening train from Cardiff. She then has to catch a connecting train.

She allows 10 minutes to change platforms at the station.

She requires the minimum probability of arriving at the platform to catch the connecting train to be 99%.

(d) Find the earliest time, correct to the nearest minute, that the connecting train could depart to meet Janine's requirements.

[4]

David, an employee of the train company, is concerned about the number of trains being late.

He spends 7 days in London noting the arrivals of trains.

It can be assumed that train arrivals are independent.

(e) Find the probability that the Cardiff to London evening train arrives late

(i) on the first day that David is in London.

(ii) on 5 of the 7 days that David is in London.

(iii) on at least 5 consecutive days that David is in London.

[9]

Turn over

2. [Maximum mark: 17]

A company produces electronic components on a large scale. They carry out quality control tests to determine whether the components meet the company's standards.

Zaakir, the owner of the company, wants the quality control team to analyse the distribution of the weights of the components.

Based on historical data, the quality control team knows that the weights of the components follow a normal distribution with a mean of 2.5 grams and a standard deviation of 0.15 grams.

- (a) Find the probability that the weight of a component selected at random is greater than 2.8 grams. [2]

The probability that the weight of a component selected at random is greater than w grams is 0.8.

- (b) (i) Sketch a diagram of a normal curve to show the area represented by this probability.
(ii) Find the value of w . [4]

To pass Test 1, the weight of a component must be between 2.3 grams and 2.7 grams.

- (c) (i) Find the probability that a randomly selected component passes Test 1.
(ii) Find the expected number of components in a box of 200 that will pass Test 1. [3]

Zaakir asks the quality control team to conduct a more in-depth analysis by performing a new test, Test 2. The probability of a component passing Test 2 is 0.95. The team randomly selects one box and tests each of the 200 components.

- (d) Find the probability that exactly 190 components pass Test 2. [2]
(e) Find the probability that at least 188 components pass Test 2. [2]

Instead of testing all 200 components, Zaakir now decides to test a random sample of 12 components from a box of 200 components. He decides that the box will only be dispatched if at least 10 of the 12 components pass both Test 1 and Test 2. The results of Test 1 and Test 2 are independent.

- (f) Find the probability that the box is dispatched. [4]

Turn over

4. [Maximum mark: 7]

Under controlled driving conditions, Jacob investigated the fuel efficiency of his car when using premium fuel compared to standard fuel.

Jacob recorded the distance travelled per litre (km L^{-1}) using standard fuel for six days and then using premium fuel for seven days. This information is shown in the following tables.

	Standard fuel						
Distance travelled per litre (km L^{-1})	8.1	8.2	7.9	8.2	7.9	8.0	

	Premium fuel						
Distance travelled per litre (km L^{-1})	8.3	8.4	8.1	8.3	8.2	8.0	8.2

At the 5% significance level, Jacob performs a t -test to determine whether there is sufficient evidence that his car travels further using premium fuel compared to standard fuel.

- (a) State one mathematical assumption made for this test to be valid.

[1]
- (b) Write down the

(i) null hypothesis;

(ii) alternative hypothesis.

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

(ii) State your conclusion to the test in context. Justify your answer.

[4]

(This question continues on the following page)



(Question 4 continued)

[illegible]

Turn over

Please **do not** write on this page.

Answers written on this page
will not be marked.



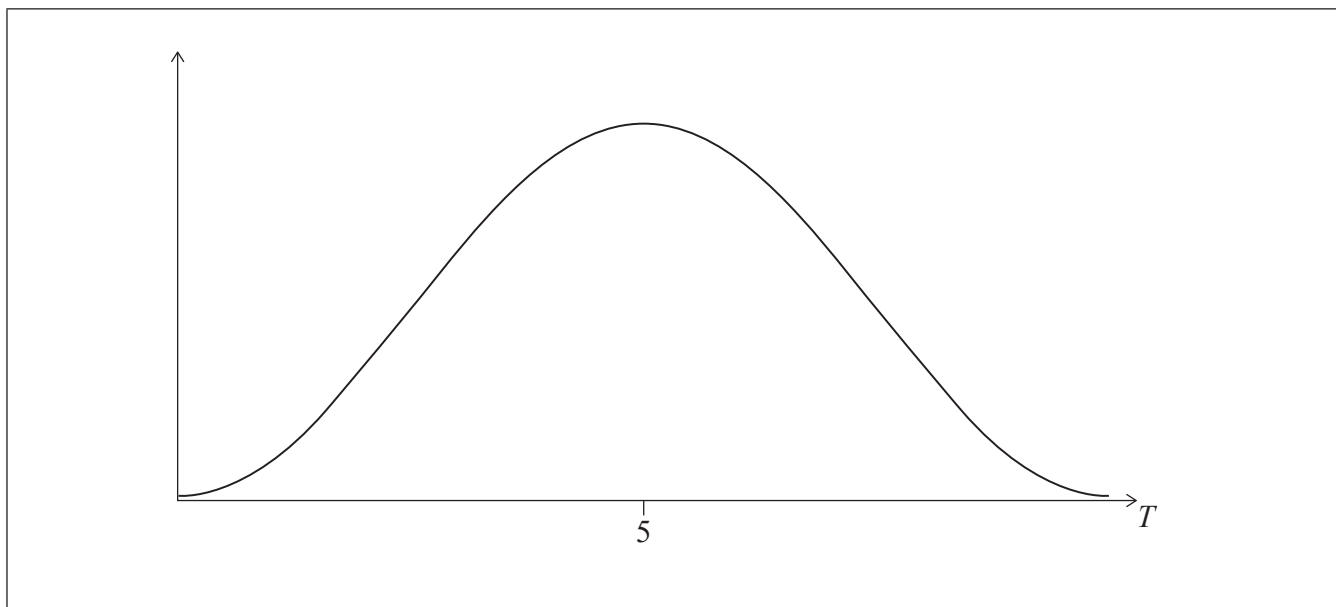
7. [Maximum mark: 5]

In a reaction test, the time, T seconds, taken to complete the test can be modelled by a normal distribution with a mean of 5 seconds and a standard deviation of 1 second.

- (a) Find the probability that a randomly selected person completes the reaction test in less than 4 seconds.

[2]

The graph of the distribution of T is shown in the following diagram.



The slowest 15% of participants will be given training to reduce their time to complete the test.

- (b) (i) Shade the region on the given diagram that corresponds to these participants.
- (ii) Find the least time taken to complete the test by a person who will be given training.

[3]

(This question continues on the following page)



(Question 7 continued)

[illegible]

Turn over

Please **do not** write on this page.

Answers written on this page
will not be marked.



9. [Maximum mark: 7]

On a television quiz show, a contestant is asked **ten questions**. Each question is multiple choice with four possible options, only one of which is correct.

Xiaoyi is a contestant and randomly guesses each answer. Her answers to the questions are independent of each other.

(a) Find the probability that Xiaoyi

- (i) answers exactly five questions correctly;
- (ii) answers at least three questions correctly.

[4]

Xiaoyi answers all ten questions correctly and is offered the choice of two prizes:

- **Option A:** \$50 000 in cash
- **Option B:** Toss a biased two-sided coin where the probability of showing heads is 0.3 and the probability of showing tails is 0.7. If the outcome is heads, Xiaoyi will win \$500 000. If the outcome is tails, Xiaoyi will receive no money.

(b) Show that the expected value of Xiaoyi's winnings if she chooses **Option B** is \$150 000.

[2]

(c) State, giving a reason, whether Xiaoyi should choose **Option A** or **Option B**.

[1]

(This question continues on the following page)



(Question 9 continued)

[illegible]

Turn over

Consider the following list of 10 data items.

This data has a mean of 19.5.

(a) mode. [1]

(b) median. [1]

(c) standard deviation. [1]

(d) range. [2]

[illegible]

3. [Maximum mark: 6]

Abigail is planning an end-of-school celebration. She asks all her classmates to choose both a main course and an ice cream flavour.

She organizes the replies from her classmates into the following table.

		Main course			
		Burgers	Chicken skewers	Salad	Pasta
Ice cream flavour	Chocolate	15	10	12	16
	Strawberry	11	14	13	12
	Vanilla	10	21	11	5

- (a) Find the total number of classmates who replied. [1]

Abigail performs a χ^2 test for independence at the 5% significance level to see if the main course choice is independent of the ice cream flavour choice.

- (b) Write down the null hypothesis. [1]
- (c) Find the p -value of Abigail's test. [2]
- (d) Write down the conclusion of the test in context. Justify your answer. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



4. [Maximum mark: 5]

The heights of pine trees in a particular forest are normally distributed, with a mean of 35 metres and a standard deviation of 5.5 metres.

- (a) Write down the probability that a randomly chosen pine tree has a height of less than 35 metres.

[1]
- (b) Find the probability that a randomly chosen pine tree has a height of greater than 25 metres.

[2]
- (c) In the forest, 8% of pine trees have a height of less than k metres.
Find the value of k .

[2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

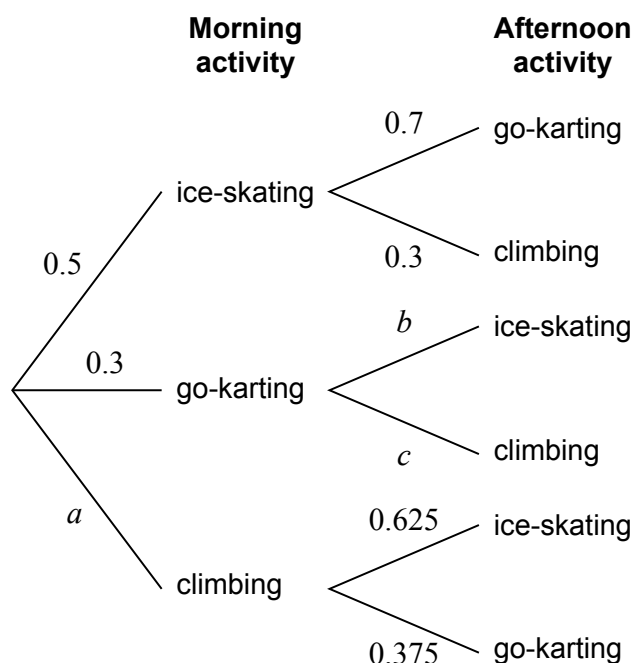
.....



Turn over

7. [Maximum mark: 7]

At a sports centre, students choose one activity in the morning and a second, different activity in the afternoon. The tree diagram shows the probability of students choosing each morning and afternoon activity.



- (a) Write down the value of a . [1]
- (b) Find the probability of a student choosing ice-skating in the morning and climbing in the afternoon. [2]
- (c) Find an expression, in terms of b , for the probability of a student choosing go-karting in the morning and ice-skating in the afternoon. [1]
- (d) At the end of the day, 87.5% of the students chose ice-skating. Calculate the value of
- (i) b . [3]
- (ii) c .

(This question continues on the following page)



(Question 7 continued)

A series of horizontal dotted lines for writing.



Turn over

10. [Maximum mark: 7]

In a job assessment there are 15 multiple choice questions, each with 5 answers. Only one answer is correct for each question.

Adil completes the assessment by randomly choosing every answer. All answers are equally likely to be chosen. Each of Adil’s choices are independent.

- (a) Write down the expected number of correct answers. [2]
- (b) Find the probability that Adil correctly answers exactly 3 questions. [3]

If Adil answers 7 or more questions correctly, he will be offered a job.

- (c) Find the probability that Adil is offered a job. [2]

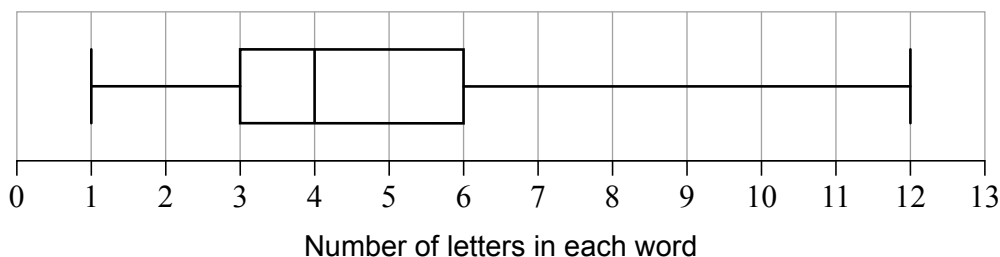


Turn over

1. [Maximum mark: 15]

Dr Petrillo wrote a short scientific essay. He analysed the readability of his essay by counting the number of letters in each word.

Dr Petrillo constructs a box and whisker diagram for his data.



(a) Write down

- (i) the median;
- (ii) the upper quartile, Q_3 ;
- (iii) the interquartile range, IQR.

[3]

Dr Petrillo now wants to modify his diagram to show any outliers. He considers the longer words in his data and uses the following formula:

$$\text{outliers} > (1.5 \times \text{IQR}) + Q_3.$$

Words with at least k letters are considered outliers.

(b) Find the value of k .

[2]

(This question continues on the following page)

(Question 1 continued)

Dr Petrillo further considers the outliers and sees no reason to exclude them from his analysis.

The length of each word in the essay, n , and its associated frequency are given in the following table.

n	$1 \leq n \leq 2$	$3 \leq n \leq 4$	$5 \leq n \leq 6$	$7 \leq n \leq 8$	$9 \leq n \leq 10$	$11 \leq n \leq 12$	$n \geq 13$
Frequency	289	480	220	136	90	45	0

- (c) Use the mid-interval values to calculate an estimate of the mean number of letters in each word. [3]

Dr Petrillo conducts a χ^2 goodness of fit test at the 1 % significance level, to test the following null hypothesis:

H_0 : The frequency of the number of letters in each word in his essay is consistent with the English language.

- (d) Write down the alternative hypothesis for this test. [1]

The observed and expected frequencies of the number of letters in each word in his essay are listed in the following table.

n	$1 \leq n \leq 2$	$3 \leq n \leq 4$	$5 \leq n \leq 6$	$7 \leq n \leq 8$	$9 \leq n \leq 10$	$11 \leq n \leq 12$	$n \geq 13$
Observed frequency	289	480	220	136	90	45	0
Expected frequency	252.0	491.4	239.4	151.2	88.2	31.8	6.0

- (e) (i) Write down the number of degrees of freedom.
 (ii) Find the χ^2 statistic for this test.
 (iii) Find the p -value for this test. [4]

The critical value for this test, at the 1 % significance level, is 16.812.

- (f) State whether Dr Petrillo should reject the null hypothesis. Justify your answer. [2]

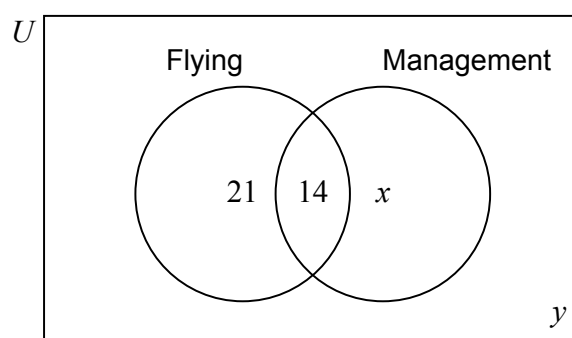
Turn over

3. [Maximum mark: 15]

Hawksburn Aviation School offers three courses, flying, management and engineering. The following information was collected about the 50 students enrolled in courses at the school.

- 21 students are enrolled in the flying course only.
- 14 students are enrolled in both the flying course and the management course.
- Students enrolled in the engineering course are not enrolled in any other courses.

The information about all 50 students is presented in the following Venn diagram.



A student is selected randomly from this group.

- (a) Find the probability that
- the student is enrolled in the flying course;
 - the student is enrolled in the management course, given that they are enrolled in the flying course.

[4]

Let F be the event that the student selected is enrolled in the flying course and M be the event that the student is enrolled in the management course.

- (b) State whether F and M are mutually exclusive. Justify your answer.

[2]

It is known that $P(F|M) = 0.7$.

- (c)
- Find the value of x .
 - Find the value of y .
 - Write down the number of students enrolled in only one course.

[6]

Three students are randomly selected from the group of 50 to give a presentation.

- (d) Find the probability that all three students are enrolled in the flying course.

[3]

Turn over

2. [Maximum mark: 14]

Aiden works for WoodCraft Wonders, a company that makes wooden letters. He wants to predict the demand for different letters. He thinks that letters earlier in the alphabet are more popular.

Aiden finds that, in the year he was born, his name was the most popular in the United States. From the 10 most popular boys' names, his name was first alphabetically. This data is shown in the following table.

Name	Popularity	Position in the alphabet
Aiden	1	1
Ethan	2	3
Jacob	3	6
Jayden	4	7
Caden	5	2
Noah	6	9
Jackson	7	5
Jack	8	4
Logan	9	8
Ryan	10	10

Aiden calculates the Spearman's rank correlation coefficient, r_s , for this data.

(a) Find r_s , and use the following table to interpret its value.

[4]

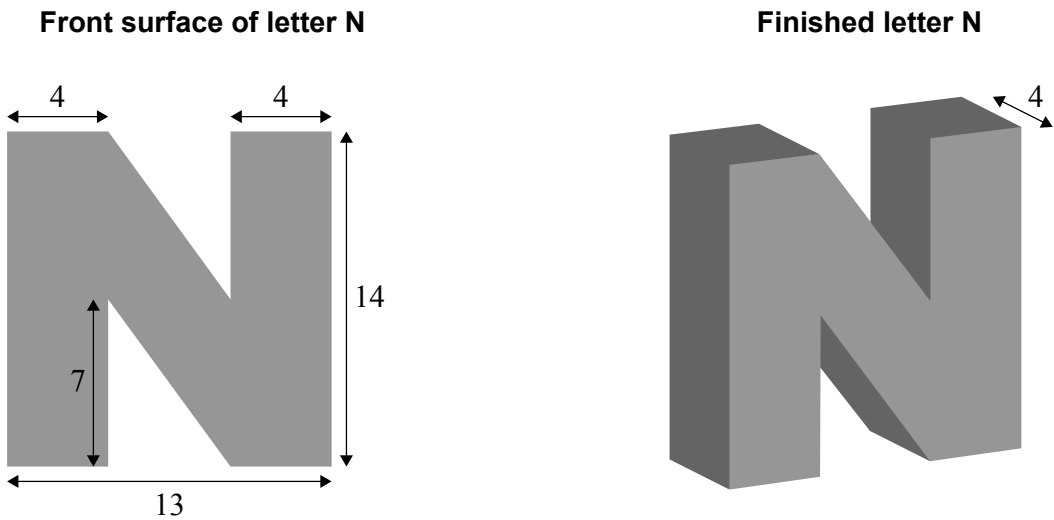
$ r_s $	Strength
0.000 to 0.199	Very weak
0.200 to 0.399	Weak
0.400 to 0.599	Moderate
0.600 to 0.799	Strong
0.800 to 1.000	Very strong

(This question continues on the following page)

(Question 2 continued)

WoodCraft Wonders makes the letter N from a wooden rectangular prism with a depth of 4 cm. The front surface is a rectangle with a base length of 13 cm and a height of 14 cm. Two identical right-angled triangular prisms are then cut out of the rectangular prism to make the letter N, as shown in the diagram.

diagrams not to scale



- (b) Find the **area** of the front surface of the letter N.

[3]
- (c) Calculate the **volume** of the finished letter N.

[1]

WoodCraft Wonders has two machines that make the wooden letters: Machine A and Machine B. Aiden wants to test if the machines are producing letters of different weights. A sample of wooden letters in the shape of an N are taken from both machines and weighed.

Machine A sample (grams)						
355	342	361	343	328	376	364

Machine B sample (grams)						
338	343	332	358	349	332	324

Aiden assumes the samples came from normally distributed populations with equal variances, and he performs a *t*-test at the 5% significance level.

- (d) (i) Write down the null hypothesis and the alternative hypothesis.

(ii) Find the *p*-value for this test.

(iii) Write down the conclusion of the test. Give a reason for your answer.

[6]

Turn over