Subject – Math AI(Higher Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2024
Paper -2
Questions

Question 1

[Maximum mark: 12]

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) 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]

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.

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

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

Colour	Brown	Red	Green	Orange	Yellow	Purple
Expected Frequency						

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

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

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

Question 2

[Maximum mark: 13]

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

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

(i) in less than $6.5\,\text{m}$.

(ii) in more than $7\,\text{m}$. [3]

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

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

(i) $6.5\,\text{m}$ and $6.75\,\text{m}$.

(ii) $6.75\,\text{m}$ and $7\,\text{m}$. [3]

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

Measured stopping distance	Number of bicycles
Less than $6.5\,\text{m}$	12
Between $6.5\,\text{m}$ and $6.75\,\text{m}$	428
Between $6.75\,\text{m}$ and $7\,\text{m}$	527
More than $7\,\text{m}$	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 $20\,\text{km}\,\text{h}^{-1}$ can be modelled by a normal distribution with mean $6.76\,\text{m}$ and standard deviation $0.12\,\text{m}$.

(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]

Question 3

[Maximum mark: 14]

Hank sets up a bird table in his garden to provide the local birds with some food. Hank notices that a specific bird, a large magpie, visits several times per month and he names him Bill. Hank models the number of times per month that Bill visits his garden as a Poisson distribution with mean 3.1.

(a) Using Hank's model, find the probability that Bill visits the garden on exactly four occasions during one particular month. [1]

(b) Over the course of 3 consecutive months, find the probability that Bill visits the garden:

(i) on exactly 12 occasions.

(ii) during the first and third month only. [5]

(c) Find the probability that over a 12-month period, there will be exactly 3 months when Bill does not visit the garden. [4]

After the first year, a number of baby magpies start to visit Hank's garden. It may be assumed that each of these baby magpies visits the garden randomly and independently, and that the number of times each baby magpie visits the garden per month is modelled by a Poisson distribution with mean 2.1.

(d) Determine the least number of magpies required, including Bill, in order that the probability of Hank's garden having at least 30 magpie visits per month is greater than 0.2. [4]

Question 4

[Maximum mark: 16]

It is known that the weights of male Persian cats are normally distributed with mean $6.1\,\text{kg}$ and variance $0.5^2\,\text{kg}^2$.

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

(b) Find the proportion of male Persian cats weighing between $5.5\,\text{kg}$ and $6.5\,\text{kg}$. [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 $5.3\,\text{kg}$. [3]

The male cats are now joined by 80 female Persian cats. The female cats are drawn from a population whose weights are normally distributed with mean $4.5\,\text{kg}$ and standard deviation $0.45\,\text{kg}$.

(d) Ten female cats are chosen at random.

(i) Find the probability that exactly one of them weighs over $4.62\,\text{kg}$.

(ii) Let N be the number of cats weighing over $4.62\,\text{kg}$.

Find the variance of N. [5]

A cat is selected at random from all 160 cats.

(e) Find the probability that the cat was female, given that its weight was over $4.7\,\text{kg}$. [4]

Question 5

[Maximum mark: 14]

Loreto is a manager at the Da Vinci health centre. If the mean rate of patients arriving at the health centre exceeds 1.5 per minute then Loreto will employ extra staff. It is assumed that the number of patients arriving in any given time period follows a Poisson distribution.

Loreto performs a hypothesis test to determine whether she should employ extra staff. She finds that 320 patients arrived during a randomly selected 3-hour clinic.

(a) (i) Write down null and alternative hypotheses for Loreto's test.

(ii) Using the data from Loreto's sample, perform the hypothesis test at a 5% significance level to determine if Loreto should employ extra staff. [7]

Loreto is also concerned about the average waiting time for patients to see a nurse. The health centre aims for at least 95% of patients to see a nurse in under 20 minutes.

Loreto assumes that the waiting times for patients are independent of each other and decides to perform a hypothesis test at a 10% significance level to determine whether the health centre is meeting its target.

Loreto surveys 150 patients and finds that 11 of them waited more than 20 minutes.

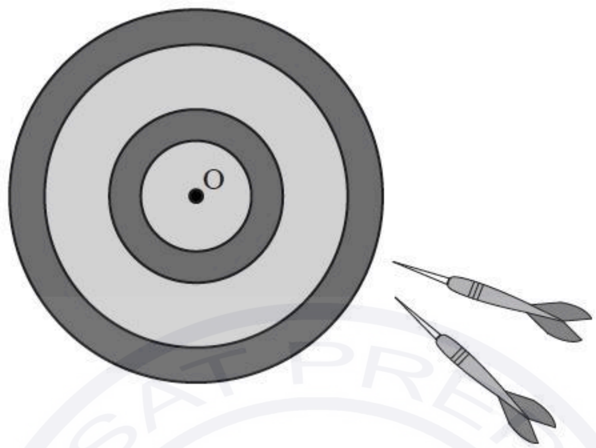
(b) (i) Write down null and alternative hypotheses for this test.

(ii) Perform the test, clearly stating the conclusion in context. [7]

Question 6

[Maximum mark: 16]

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 Arianne's expected score in the competition.

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

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

(iv) Given that Arianne scores at least 5 points, find the probability that Arianne scores less than 8 points. [9]

Question 7

[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]

Question 8

[Maximum mark: 13]

A Principal would like to compare the students in his school with a national standard. He decides to give a test to eight students made up of four boys and four girls. One of the teachers offers to find the volunteers from his class.

(a) Name the type of sampling that best describes the method used by the Principal. [1]

The marks out of 40, for the students who took the test, are:

$$25,\ 29,\ 38,\ 37,\ 12,\ 18,\ 27,\ 31.$$

(b) For the eight students find

(i) the mean mark.

(ii) the standard deviation of the marks. [3]

The national standard mark is 25.2 out of 40.

(c) Perform an appropriate test at the 5% significance level to see if the mean marks achieved by the students in the school are higher than the national standard. It can be assumed that the marks come from a normal population. [5]

(d) State one reason why the test might not be valid. [1]

Two additional students take the test at a later date and the mean mark for all ten students is 28.1 and the standard deviation is 8.4.

For further analysis, a standardized score out of 100 for the ten students is obtained by multiplying the scores by 2 and adding 20.

(e) For the ten students, find

(i) their mean standardized score.

(ii) the standard deviation of their standardized score. [3]

Question 9

[Maximum mark: 18]

A company makes doors for kitchen cupboards from two layers. The inside layer is wood, and its thickness is normally distributed with mean 7 mm and standard deviation 0.3 mm. The outside layer is plastic, and its thickness is normally distributed with mean 3 mm and standard deviation 0.16 mm. The thickness of the plastic is independent of the thickness of the wood.

(a) Find the probability that a randomly chosen door has a total thickness of less than 9.5 mm. [5]

Eight doors are to be packed into a box to send to a customer. The width of the box is 82 mm. The thickness of each door is independent.

(b) Find the probability that the total thickness of the eight doors is greater than the width of the box. [4]

The company buys two new machines, A and B, to make the wooden layers. An employee claims that the layers from machine B are thinner than the layers from machine A. In order to test this claim, a random sample is taken from each machine.

The seven layers in the sample from machine A have a thickness, in mm, of

$$6.23,\ 7.04,\ 7.31,\ 6.79,\ 6.91,\ 6.79,\ 7.47.$$

(c) Find the

(i) mean.

(ii) unbiased estimate of the population variance. [3]

The eight layers in the sample from machine B have a mean thickness of 6.89 mm and $s_{n-1} = 0.31$.

(d) Perform a suitable test, at the 5% significance level, to test the employee's claim. You may assume the thickness of the wooden layers from each machine are normally distributed with equal population variance. [6]

Question 10

[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]

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

(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]

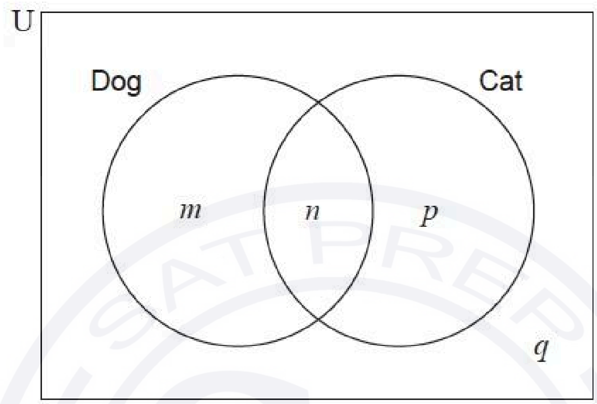
(f) When calculating the ranks, Chester incorrectly read the Netherlands' score as 478. Explain why the value of the Spearman's rank correlation r_s does not change despite this error. [1]

Question 11

[Maximum mark: 15]

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 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]

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

(c) (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.

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

Question 12

[Maximum mark: 18]

The gardener in a local park suggested that the number of snails found in the park can be modelled by a Poisson distribution.

(a) Suggest two observations that the gardener may have made that led him to suggest this model. [2]

Now assume that the model is valid and that the mean number of snails per m^2 is 0.2.
The gardener inspects, at random, a $12\,\text{m}^2$ area of the park.

(b) Find the probability that the gardener finds exactly four snails. [3]

(c) Find the probability that the gardener finds fewer than three snails. [2]

(d) Find the probability that, in three consecutive inspections, the gardener finds at least one snail per inspection. [3]

Following heavy rain overnight, the gardener wished to determine whether the number of snails found in a random $12\,\text{m}^2$ area of the park had increased.

(e) State the hypotheses for the test. [2]

(f) Find the critical region for the test at the $1\,\%$ significance level. [3]

(g) Given that the mean number of snails per m^2 has actually risen to 0.75, find the probability that the gardener makes a Type II error. [3]

Question 13

[Maximum mark: 15]

Goran is interested in the number of sightings of a particular bird each week in the 50 weeks following the first day of September. He collects some data which is shown in the table.

Number of sightings	0	1	2	3	4	5	More than 5
Number of weeks	8	16	13	8	3	2	0

The sample mean number of sightings per week for this data is 1.76.

(a) Calculate the unbiased estimate of the population variance of sightings per week. [3]

Goran believes that the data follows a Poisson distribution.

(b) State why your answer to part (a) supports Goran's belief. [1]

Goran decides to test at the 5% significance level to see if his belief is correct.

His null hypothesis is $X \sim \text{Po}(m)$, where the random variable, X, is defined as the number of sightings per week.

Goran estimates parameter m to be the mean of the sample, 1.76. He calculates the expected frequencies for sightings per week in the 50 weeks after the first day of September. These are shown to two decimal places in the following table.

Number of sightings	0	1	2	3	4	5 or more
Expected frequencies	8.60	15.14	13.32	7.82	j	k

(c) Find the value of

(i) j;

(ii) k. [5]

(d) State a reason why Goran should combine groups to conduct his significance test. [1]

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

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

(g) State the conclusion of the test. Justify your answer. [2]

Question 14

[Maximum mark: 17]

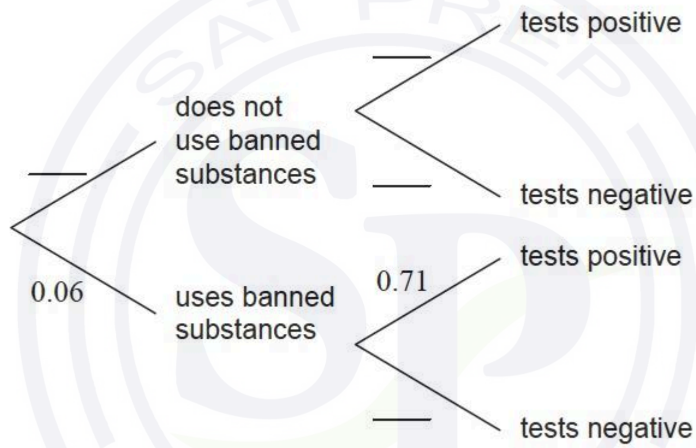
A large international sports tournament tests their athletes for banned substances.
They interpret a positive test result as meaning that the athlete uses banned substances.
A negative result means that they do not.

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]

Question 15

[Maximum mark: 13]

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 °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]

Question 16

[Maximum mark: 17]

The city of Melba has an adult population of four million people. It is assumed that the weights of adults in Melba can be modelled by a normal distribution with mean $72\,\text{kg}$ and standard deviation $10\,\text{kg}$.

(a) If 10 adults in Melba are chosen independently and at random, find the probability that more than 3 of them have a weight greater than $85\,\text{kg}$. [4]

Laetitia runs a travel agency in Melba. The elevator to her office is designed to hold a maximum of 8 people.

(b) Write down a probability distribution that models the total weight of 8 adults chosen independently and at random from Melba. [3]

The total weight of 8 adults exceeds w on less than 1% of all occasions that 8 adults enter the elevator.

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

A newspaper claims that 42% of the adults in Melba who go on holiday choose to go abroad. Laetitia believes that this is an overestimation of the true number. During the past month, Laetitia found that 67 of her clients chose a holiday abroad, and 133 chose a holiday that was not abroad.

(d) Laetitia decides to perform a test using the binomial distribution on her data for the population proportion, p, that go on holiday abroad.

(i) State **two** assumptions that Laetitia makes in order to conduct the test.

(ii) Write down the null and the alternative hypotheses for Laetitia's test, in terms of p.

(iii) Using the data from Laetitia's sample, perform the test at a 5% significance level to determine whether Laetitia's belief is reasonable. [8]

Question 17

2. [Maximum mark: 18]

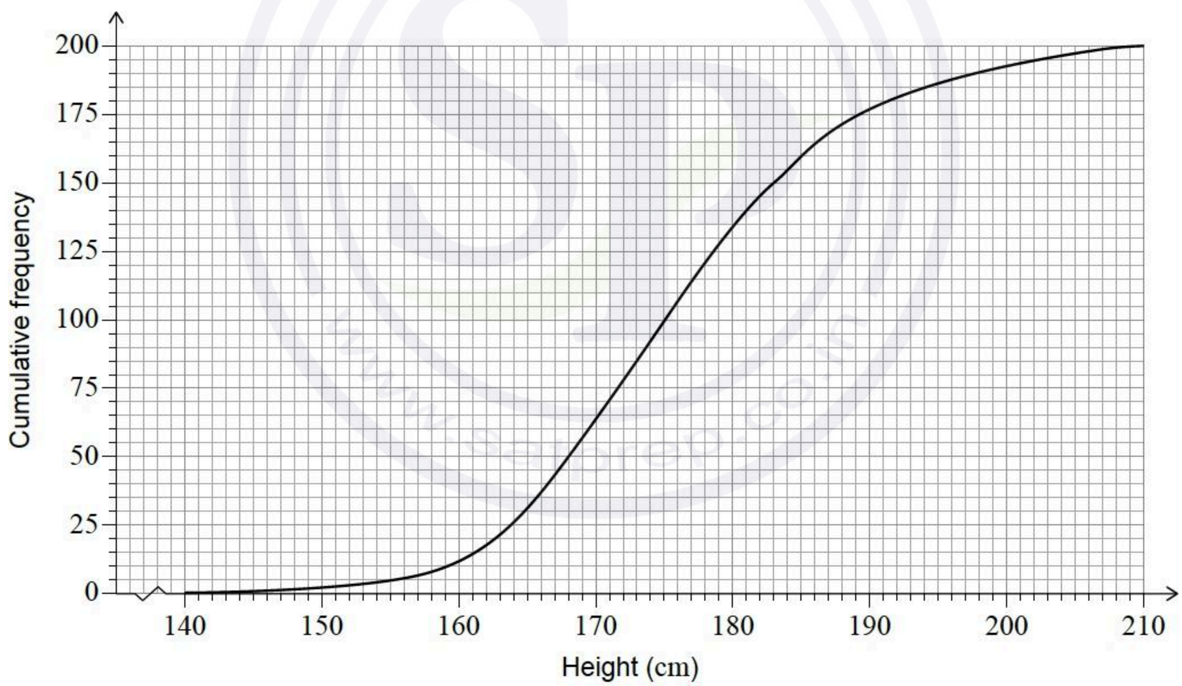
The heights, h, of 200 university students are recorded in the following table.

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

(a) (i) Write down the mid-interval value of $140 \le h < 160$.

(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]

It is believed that the heights of university students follow a normal distribution with mean $176\,\text{cm}$ and standard deviation $13.5\,\text{cm}$.

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
$160 \le h < 170$	51	42.1
$170 \le h < 180$	68	a
$180 \le h < 190$	47	46.7
$190 \le h$	23	b

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

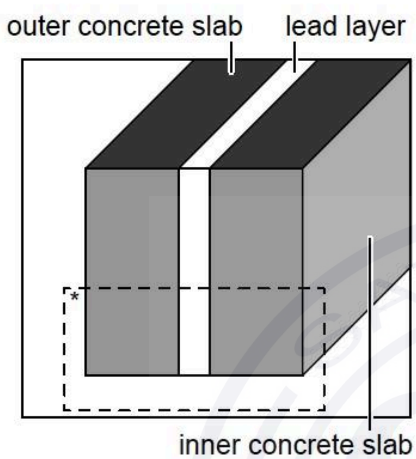
(ii) Hence, perform the test to a 5% significance level, clearly stating the conclusion in context. [8]

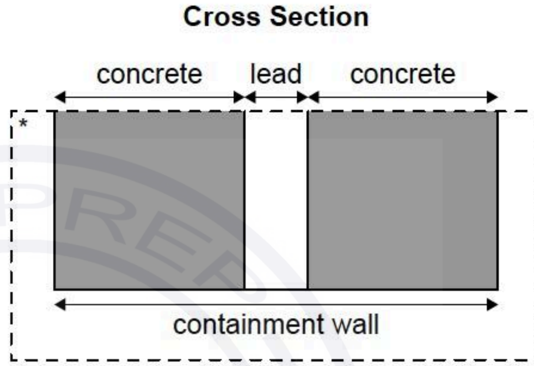
Question 18

[Maximum mark: 17]

Containment walls to protect against radiation are constructed from two parallel concrete slabs that have a layer of lead between them as shown in the diagram.

diagram not to scale





The width of a concrete slab is modelled by a normal distribution with mean $350\,\text{mm}$ and standard deviation $10\,\text{mm}$.

(a) Find the probability that a randomly chosen concrete slab is less than $340\,\text{mm}$ in width. [2]

(b) Find the endpoints of the interval, symmetric about the mean, such that $95\,\%$ of the slabs have a width that lies in this interval. [3]

Stephen assumes the lead layer is also modelled by a normal distribution, but with mean $100\,\text{mm}$ and standard deviation $5\,\text{mm}$ and is independent of the width of the slabs.

Let W be the random variable that represents the total width of the wall, measured in mm.

(c) (i) Given that the widths of any two concrete slabs are independent, calculate Stephen's value for the mean and standard deviation of W.

(ii) Hence find $\text{P}(780 < W < 810)$. [7]

There are concerns that the mean and standard deviation for Stephen's model of the lead layer are incorrect. However, his assumption that the model is normal and the width of the lead is independent of the width of the concrete slabs still holds.

On investigation it is found that the total width of the containment wall is normally distributed with mean $810\,\text{mm}$ and standard deviation $16\,\text{mm}$. The model for the width of a concrete slab does not change.

(d) Use the results for the **sum** of independent random variables to find a revised value for

(i) the mean of the width of the lead layer.

(ii) the standard deviation of the width of the lead layer. [4]

Under this revised model, $80\,\%$ of the lead layers have a width less than $k\,\text{mm}$.

(e) Calculate the value of k. [1]

Question 19

[Maximum mark: 14]

The $k\text{th}$ triangle number, T_k, is defined as $T_k = \sum_{r=1}^{k} r$.

(a) (i) Calculate the value of the fifth triangle number, T_5.

(ii) Determine the formula for T_k in the form $ak^2 + bk$. [4]

(b) (i) Find the value of $T_5 + T_4$.

(ii) Find the simplest expression for $T_k + T_{k-1}$. [3]

A bag contains 15 red discs and 10 blue discs, all identical except for colour. Two discs are chosen at random from the bag without replacement.

(c) Calculate the probability that the two discs are different colours. [3]

A bag contains T_k red discs and T_{k-1} blue discs, all identical except for colour.
Two discs are chosen at random from the bag without replacement.

(d) Show that the probability that the two discs are different colours is independent of k. [4]

Question 20

[Maximum mark: 13]

Taylor is playing a computer game in which they shoot at spaceships and battleships. The number of spaceships they hit per minute can be modelled by a Poisson distribution with mean 4.2. The number of battleships they hit per minute can be modelled by a Poisson distribution with a mean of 2.3. Any single hit occurs independently of all others.

(a) Find the probability Taylor hits

(i) at most 10 spaceships in 2 minutes.

(ii) a total of more than 10 spaceships and battleships in one minute. [5]

Every spaceship that is hit earns Taylor 3 points and every battleship 5 points. Let T be the total points earned in one minute.

(b) Find

(i) $\mathrm{E}(T)$.

(ii) $\mathrm{Var}(T)$. [3]

(c) State one reason why the distribution of T cannot be Poisson. [1]

Taylor intends to play the game for one hour.

(d) Use the central limit theorem to find the probability that Taylor's mean score per minute is greater than 25. [4]

Question 21

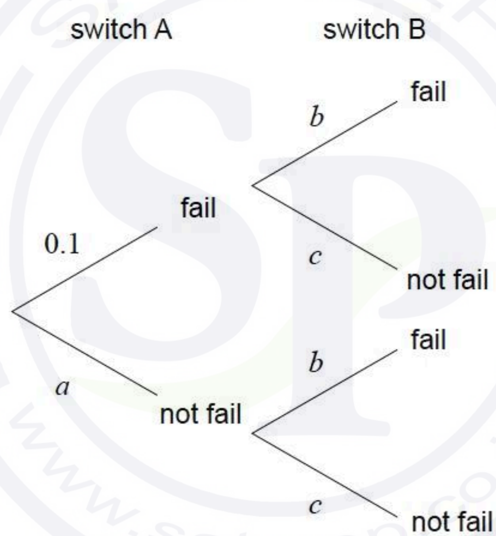
[Maximum mark: 12]

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]

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) 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. [9]

Question 22

[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 $\mathrm{P}(T < 64)$. [1]

(b) Find $\mathrm{P}(44 < T < 64)$. [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 $\mathrm{P}(T < k)$.

(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 $1\,\mathrm{kg}$ or less. Heavier packages are charged an additional fee of $\$2.00$ per kg. This fee is applied for any weight **in excess of** $1\,\mathrm{kg}$. For example, a $1.5\,\mathrm{kg}$ package is charged an additional $\$1.00$.

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

(f) Find the amount Fred Express charges for a $5.3\,\mathrm{kg}$ package. [1]

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

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

Question 23

[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]

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]