Subject – Math AI(Higher Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2022
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 $20\,\text{km h}^{-1}$ 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 $20\,\text{km h}^{-1}$ 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 $20\,\text{km h}^{-1}$ 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 $20\,\text{km h}^{-1}$ 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.

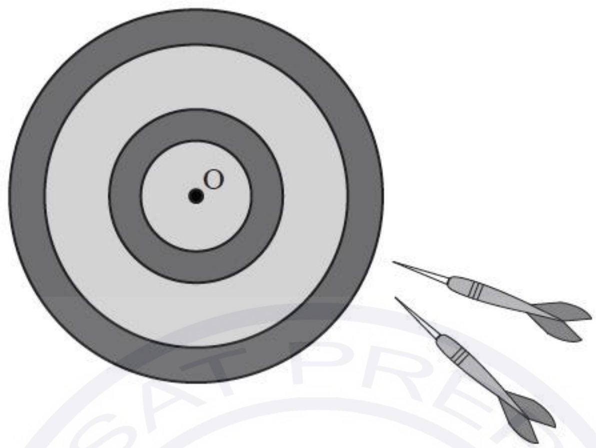
(b) (i) Write down null and alternative hypotheses for this test.

(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 $10\,\text{cm}$ and standard deviation $3\,\text{cm}$.

(a) Find the probability that

(i) a dart lands less than $13\,\text{cm}$ from O.

(ii) a dart lands more than $15\,\text{cm}$ 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 $15\,\text{cm}$ 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: 18]

A company makes doors for kitchen cupboards from two layers. The inside layer is wood, and its thickness is normally distributed with mean $7\,\text{mm}$ and standard deviation $0.3\,\text{mm}$. The outside layer is plastic, and its thickness is normally distributed with mean $3\,\text{mm}$ and standard deviation $0.16\,\text{mm}$. 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 $9.5\,\text{mm}$. [5]

Eight doors are to be packed into a box to send to a customer. The width of the box is $82\,\text{mm}$. 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 $6.89\,\text{mm}$ 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 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: 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]

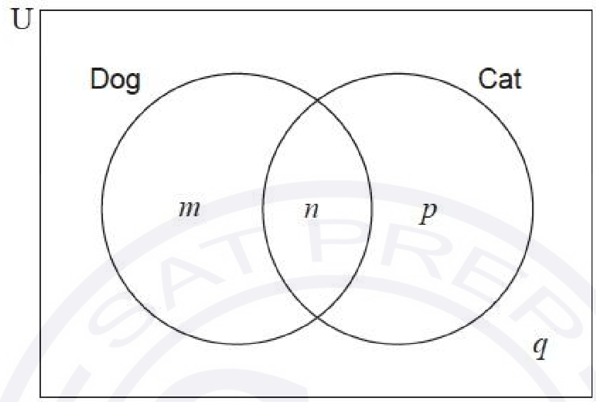
(f) When calculating the ranks, Chester incorrectly read the Netherlands' score as 478. Explain why the value of the Spearman's rank correlation r_s does not change despite this error. [1]

Question 10

[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]