Subject – Math AI(Standard Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2022
Paper -2
Questions

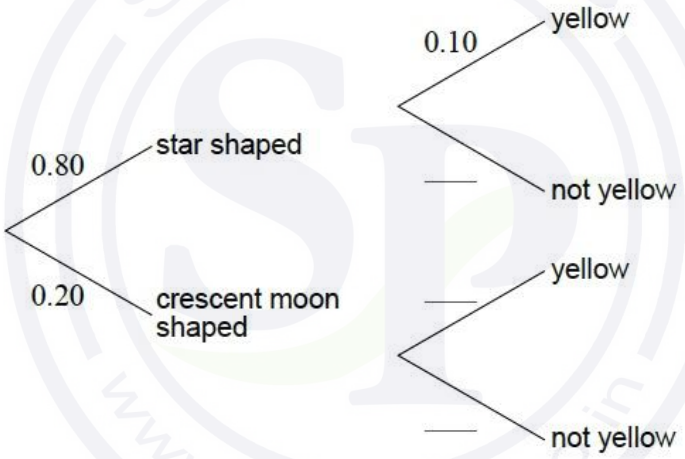
Question 1

[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]

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]

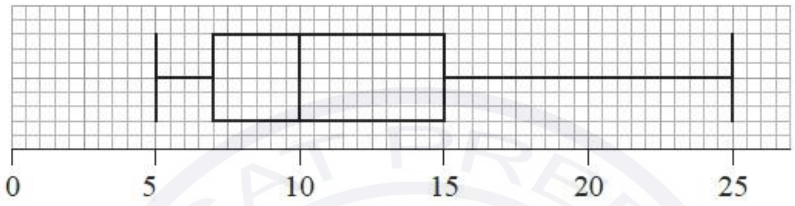
Question 3

[Maximum mark: 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.



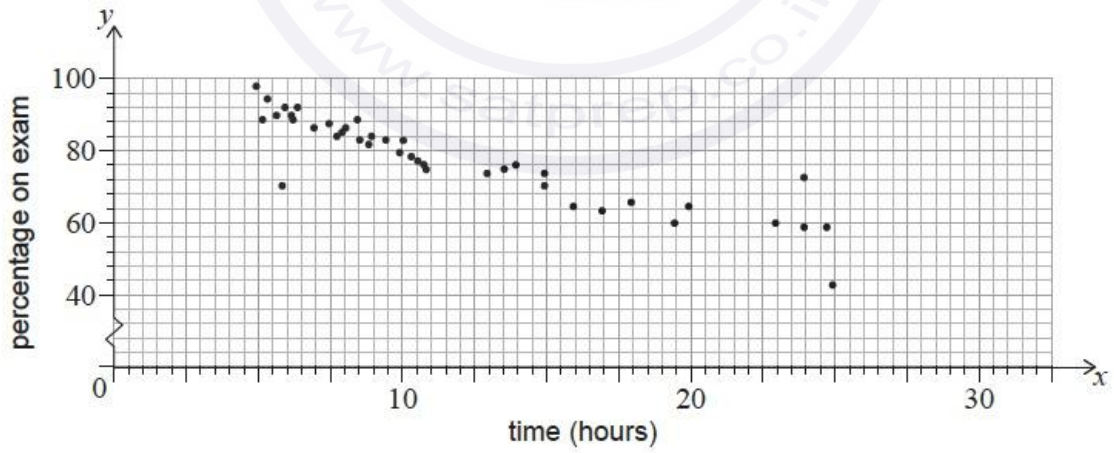
(b) Write down the median time to read the book. [1]

(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]

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.

(ii) Interpret your result. [3]

Question 4

[Maximum mark: 14]

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]

(d) It is found that 12 of the cats weigh more than $x\,\text{kg}$. 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 $6.25\,\text{kg}$. [4]

Question 5

[Maximum mark: 19]

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

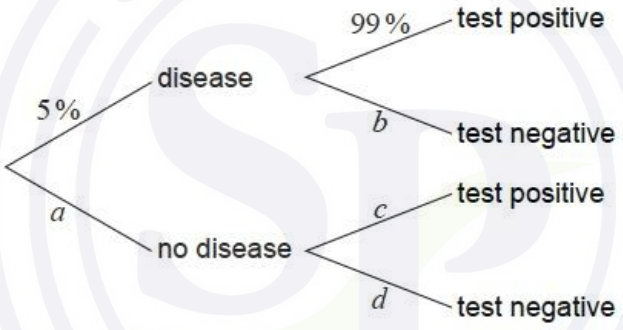
They test every patient who comes to the centre on a particular day.

(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]

(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.

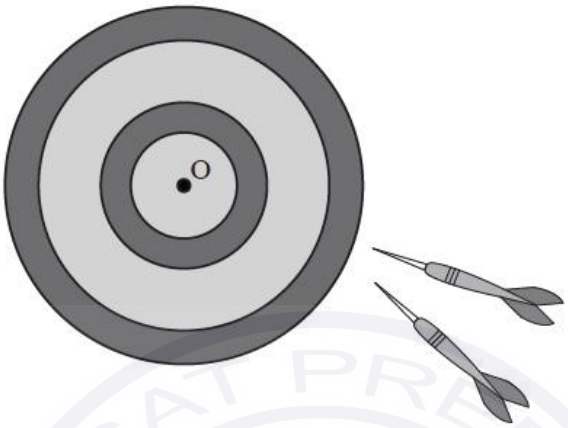
(e) Draw a Venn diagram to illustrate this information, placing all relevant information on the diagram. [3]

(f) Find the total number of patients who visited the centre during this day. [2]

Question 6

[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 $10\,\text{cm}$ 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 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]

Question 7

[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]

Question 8

[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 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.

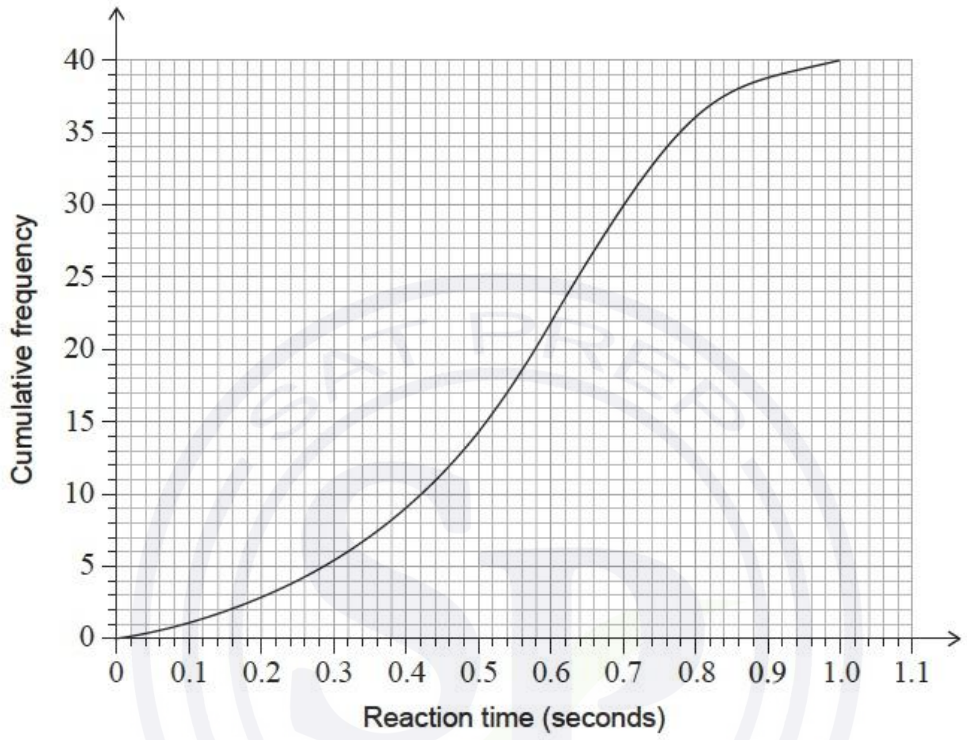
(ii) Interpret the value obtained for r_s. [3]

(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: 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]

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 $0.2 < t \leq 0.4$.

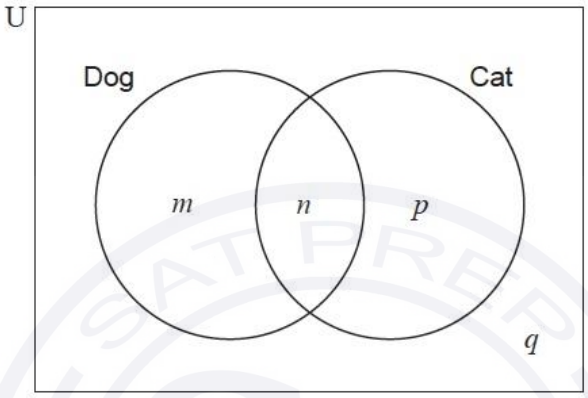
(g) Suggest how, if at all, the estimated mean and estimated median reaction times will change if the errors are corrected. Justify your response. [4]

Question 10

[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]

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]

Question 11

[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]

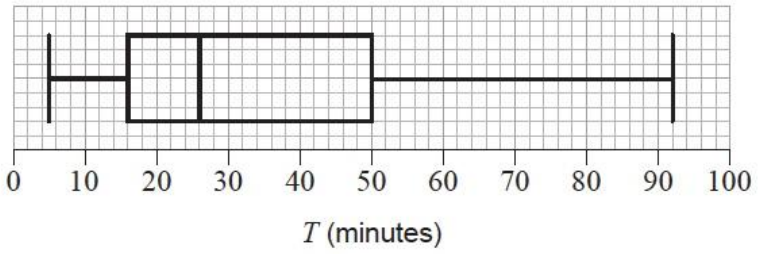
(d) Use Elsie's data to calculate an estimate of the mean time that people spent in the library. [2]

(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]

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]