Subject – Math AI(Higher Level)
Topic - Statistics and Probability
Year - May 2021 – Nov 2022
Paper -3
Questions

Question 1

[Maximum mark: 27]

Two IB schools, A and B, follow the IB Diploma Programme but have different teaching methods. A research group tested whether the different teaching methods lead to a similar final result.

For the test, a group of eight students were randomly selected from each school. Both samples were given a standardized test at the start of the course and a prediction for total IB points was made based on that test; this was then compared to their points total at the end of the course.

Previous results indicate that both the predictions from the standardized tests and the final IB points can be modelled by a normal distribution.

It can be assumed that:
- the standardized test is a valid method for predicting the final IB points
- that variations from the prediction can be explained through the circumstances of the student or school.

(a) Identify a test that might have been used to verify the null hypothesis that the predictions from the standardized test can be modelled by a normal distribution. [1]

(b) State why comparing only the final IB points of the students from the two schools would not be a valid test for the effectiveness of the two different teaching methods. [1]

The data for school A is shown in the following table.

School A

Student number	Gender	Predicted IB points (p)	Final IB points (f)
1	male	43.2	44
2	male	36.5	34
3	female	37.1	38
4	male	30.9	28
5	male	41.1	39
6	female	35.1	39
7	male	36.4	40
8	male	38.2	38
	Mean	37.31	37.5

(c) For each student, the change from the predicted points to the final points ($f-p$) was calculated.

(i) Find the mean change.

(ii) Find the standard deviation of the changes. [3]

(d) Use a paired t-test to determine whether there is significant evidence that the students in school A have improved their IB points since the start of the course. [4]

The data for school B is shown in the following table.

School B

Student number	Gender	Final IB points– Predicted IB points ($f-p$)
1	male	8.7
2	female	−1.1
3	female	4.8
4	female	−1.5
5	male	2.5
6	female	3.2
7	female	−1.3
8	female	3.1
	Mean	2.3

(e) (i) Use an appropriate test to determine whether there is evidence, at the 5% significance level, that the students in school B have improved more than those in school A.

(ii) State why it was important to test that both sets of points were normally distributed. [6]

School A also gives each student a score for effort in each subject. This effort score is based on a scale of 1 to 5 where 5 is regarded as outstanding effort.

Student number	Gender	Predicted IB points	Final IB points	Average effort score
1	male	43.2	44	4.4
2	male	36.5	34	4.2
3	female	37.1	38	4.7
4	male	30.9	28	4.3
5	male	41.1	39	3.9
6	female	35.1	39	4.9
7	male	36.4	40	4.9
8	male	38.2	38	4.3
	Mean	37.31	37.5	4.45

It is claimed that the effort put in by a student is an important factor in improving upon their predicted IB points.

(f) (i) Perform a test on the data from school A to show it is reasonable to assume a linear relationship between effort scores and improvements in IB points. You may assume effort scores follow a normal distribution.

(ii) Hence, find the expected improvement between predicted and final points for an increase of one unit in effort grades, giving your answer to one decimal place. [4]

A mathematics teacher in school A claims that the comparison between the two schools is not valid because the sample for school B contained mainly girls and that for school A, mainly boys. She believes that girls are likely to show a greater improvement from their predicted points to their final points.

She collects more data from other schools, asking them to class their results into four categories as shown in the following table.

	$(f-p)<-2$	$-2\leq(f-p)<0$	$0\leq(f-p)<2$	$(f-p)\geq 2$
Male	6	8	10	9
Female	3	8	14	8

(g) Use an appropriate test to determine whether showing an improvement is independent of gender. [6]

(h) If you were to repeat the test performed in part (e) intending to compare the quality of the teaching between the two schools, suggest **two** ways in which you might choose your sample to improve the validity of the test. [2]

Question 2

[Maximum mark: 28]

A firm wishes to review its recruitment processes. This question considers the validity and reliability of the methods used.

Every year an accountancy firm recruits new employees for a trial period of one year from a large group of applicants.

At the start, all applicants are interviewed and given a rating. Those with a rating of either *Excellent*, *Very good* or *Good* are recruited for the trial period. At the end of this period, some of the new employees will stay with the firm.

It is decided to test how valid the interview rating is as a way of predicting which of the new employees will stay with the firm.

Data is collected and recorded in a contingency table.

	Interview rating		
	Excellent	**Very good**	**Good**
Stay	12	20	19
Leave	10	15	24

(a) Use an appropriate test, at the 5% significance level, to determine whether a new employee staying with the firm is independent of their interview rating. State the null and alternative hypotheses, the p-value and the conclusion of the test. [6]

The next year's group of applicants are asked to complete a written assessment which is then analysed. From those recruited as new employees, a random sample of size 18 is selected.

The sample is stratified by department. Of the 91 new employees recruited that year, 55 were placed in the national department and 36 in the international department.

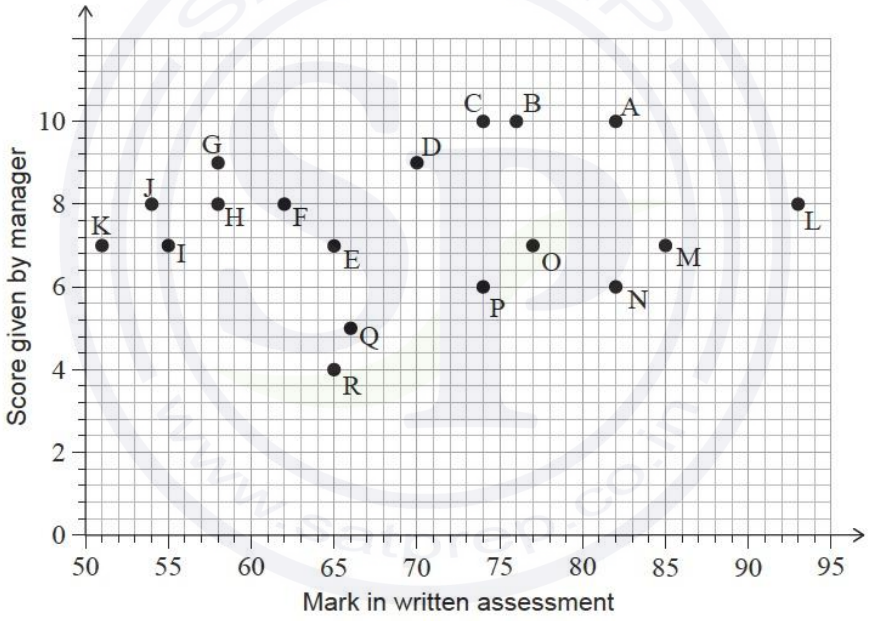
(b) Show that 11 employees are selected for the sample from the national department. [2]

(Question 2 continued)

At the end of their first year, the level of performance of each of the 18 employees in the sample is assessed by their department manager. They are awarded a score between 1 (low performance) and 10 (high performance).

The marks in the written assessment and the scores given by the managers are shown in both the table and the scatter diagram.

	National department											International department						
Employee	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Mark in written assessment	82	76	74	70	65	62	58	58	55	54	51	93	85	82	77	74	66	65
Score given by manager	10	10	10	9	7	8	9	8	7	8	7	8	7	6	7	6	5	4



The firm decides to find a Spearman's rank correlation coefficient, r_s, for this data.

(c) (i) Without calculation, explain why it might not be appropriate to calculate a correlation coefficient for the whole sample of 18 employees. [2]

(ii) Find r_s for the seven employees working in the **international** department. [4]

(iii) Hence comment on the validity of the written assessment as a measure of the level of performance of employees in this department. Justify your answer. [2]

(Question 2 continued)

The same seven employees are given the written assessment a second time, at the end of the first year, to measure its reliability. Their marks are shown in the table below.

	International department						
	L	M	N	O	P	Q	R
First mark	93	85	82	77	74	66	65
Second mark	90	92	85	73	79	71	65

(d) (i) State the name of this type of test for reliability. [1]

(ii) For the data in this table, test the null hypothesis, $H_0 : \rho = 0$, against the alternative hypothesis, $H_1 : \rho > 0$, at the 5% significance level. You may assume that all the requirements for carrying out the test have been met. [4]

(iii) Hence comment on the reliability of the written assessment. [1]

(e) The written assessment is in five sections, numbered 1 to 5. At the end of the year, the employees are also given a score for each of five professional attributes: V, W, X, Y and Z.

The firm decides to test the hypothesis that there is a correlation between the mark in a section and the score for an attribute.

They compare marks in **each** of the sections with scores for **each** of the attributes.

(i) Write down the number of tests they carry out. [1]

(ii) The tests are performed at the 5% significance level.

Assuming that:

- there is no correlation between the marks in any of the sections and scores in any of the attributes,
- the outcome of each hypothesis test is independent of the outcome of the other hypothesis tests,

find the probability that at least one of the tests will be significant. [4]

(iii) The firm obtains a significant result when comparing section 2 of the written assessment and attribute X. Interpret this result. [1]

Question 3

[Maximum mark: 24]

Juliet is a sociologist who wants to investigate if income affects happiness amongst doctors. This question asks you to review Juliet's methods and conclusions.

Juliet obtained a list of email addresses of doctors who work in her city. She contacted them and asked them to fill in an anonymous questionnaire. Participants were asked to state their annual income and to respond to a set of questions. The responses were used to determine a *happiness score* out of 100. Of the 415 doctors on the list, 11 replied.

(a) (i) Describe **one** way in which Juliet could improve the reliability of her investigation. [1]

(ii) Describe **one** criticism that can be made about the validity of Juliet's investigation. [1]

Juliet's results are summarized in the following table.

Response	Annual income ($)	Happiness score
A	65 000	60
B	63 000	52
C	40 000	31
D	125 000	81
E	100 000	48
F	245 000	61
G	48 000	42
H	39 000	40
I	85 000	57
J	92 000	53
K	123 456 789	56

(b) Juliet classifies response K as an outlier and removes it from the data. Suggest **one** possible justification for her decision to remove it. [1]

(c) For the remaining ten responses in the table, Juliet calculates the mean happiness score to be 52.5.

(i) Calculate the mean **annual income** for these remaining responses. [2]

(ii) Determine the value of r, Pearson's product-moment correlation coefficient, for these remaining responses. [2]

Juliet decides to carry out a hypothesis test on the correlation coefficient to investigate whether increased annual income is associated with greater happiness.

(d) (i) State why the hypothesis test should be one-tailed. [1]

(ii) State the null and alternative hypotheses for this test. [2]

The critical value for this test, at the 5% significance level, is 0.549. Juliet assumes that the population is bivariate normal.

(iii) Determine whether there is significant evidence of a positive correlation between annual income and happiness. Justify your answer. [2]

(e) Juliet wants to create a model to predict how changing annual income might affect happiness scores. To do this, she assumes that annual income in dollars, X, is the independent variable and the happiness score, Y, is the dependent variable.

She first considers a linear model of the form

$$Y = aX + b.$$

(i) Use Juliet's data to find the value of a and of b. [1]

(ii) Interpret, referring to income and happiness, what the value of a represents. [1]

Juliet then considers a quadratic model of the form

$$Y = cX^2 + dX + e.$$

(iii) Find the value of c, of d and of e. [1]

(iv) Find the coefficient of determination for each of the two models she considers. [2]

(v) Hence compare the two models. [1]

Juliet decides to use the coefficient of determination to choose between these two models.

(vi) Comment on the validity of her decision. [1]

After presenting the results of her investigation, a colleague questions whether Juliet's sample is representative of all doctors in the city.

A report states that the mean annual income of doctors in the city is $\$80\,000$. Juliet decides to carry out a test to determine whether her sample could realistically be taken from a population with a mean of $\$80\,000$.

(f) (i) State the name of the test which Juliet should use. [1]

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

(iii) Perform the test, using a 5% significance level, and state your conclusion in context. [3]

Question 4

[Maximum mark: 27]

This question uses statistical tests to investigate whether advertising leads to increased profits for a grocery store.

Aimmika is the manager of a grocery store in Nong Khai. She is carrying out a statistical analysis on the number of bags of rice that are sold in the store each day. She collects the following sample data by recording how many bags of rice the store sells each day over a period of 90 days.

Number of bags of rice sold	0	1	2	3	4	5	6	7	8	9	10
Number of days	1	8	12	11	19	14	13	8	2	0	2

She believes that her data follows a Poisson distribution.

(a) (i) Find the mean and variance for the sample data given in the table. [2]

(ii) Hence state why Aimmika believes her data follows a Poisson distribution. [1]

(b) State one assumption that Aimmika needs to make about the sales of bags of rice to support her belief that it follows a Poisson distribution. [1]

Aimmika knows from her historic sales records that the store sells an average of 4.2 bags of rice each day. The following table shows the expected frequency of bags of rice sold each day during the 90 day period, assuming a Poisson distribution with mean 4.2.

Number of bags of rice sold	≤ 1	2	3	4	5	6	7	≥ 8
Expected frequency	a	11.903	16.665	b	14.698	10.289	6.173	c

(c) Find the value of a, of b, and of c. Give your answers to 3 decimal places. [5]

Aimmika decides to carry out a χ^2 goodness of fit test at the 5% significance level to see whether the data follows a Poisson distribution with mean 4.2.

(d) (i) Write down the number of degrees of freedom for her test. [1]

(ii) Perform the χ^2 goodness of fit test and state, with reason, a conclusion. [7]

Aimmika claims that advertising in a local newspaper for 300 Thai Baht (THB) per day will increase the number of bags of rice sold. However, Nichakarn, the owner of the store, claims that the advertising will **not** increase the store's overall profit.

Nichakarn agrees to advertise in the newspaper for the next 60 days. During that time, Aimmika records that the store sells 282 bags of rice with a profit of 495 THB on each bag sold.

(e) Aimmika wants to carry out an appropriate hypothesis test to determine whether the number of bags of rice sold during the 60 days increased when compared with the historic sales records.

(i) By finding a critical value, perform this test at a 5% significance level. [6]

(ii) Hence state the probability of a Type I error for this test. [1]

(f) By considering the claims of both Aimmika and Nichakarn, explain whether the advertising was beneficial to the store. [3]

Question 5

[Maximum mark: 29]

In this question, you will explore possible approaches to using historical sports results for making predictions about future sports matches.

Two friends, Peter and Helen, are discussing ways of predicting the outcomes of international football matches involving Argentina.

Peter suggests analysing historical data to help make predictions. He lists the results of the most recent 240 matches in which Argentina played, in chronological order, then considers blocks of four matches at a time. He counts how many times Argentina has won in each block. The following table shows his results for the 60 blocks of four matches.

Number of wins for Argentina (per block of four matches)	Frequency
0	0
1	11
2	21
3	21
4	7

(a) Determine the mean number of wins per block of four matches for Argentina. [2]

Peter thinks that this data can be modelled by a binomial distribution with $n = 4$ and decides to carry out a χ^2 goodness of fit test.

(b) Use Peter's data to write down an estimate for the probability p for this binomial model. [1]

(c) (i) Use the binomial model to find the probability that Argentina win zero matches in a block of four matches. [1]

(ii) Find the expected frequency for zero wins. [2]

As some expected frequencies are less than 5, Peter combines rows in his table to produce the following observed frequencies. He then uses his binomial model to find appropriate expected frequencies, correct to one decimal place.

Number of wins for Argentina (per block of four matches)	Observed frequency	Expected frequency
0 or 1	11	10.8
2	21	20.7
3	21	20.7
4	7	7.8

(d) Peter uses this table to carry out a χ^2 goodness of fit test, to test the hypothesis that the data follows a binomial distribution with $n = 4$, at the 5% significance level.

For this test, state

(i) the null hypothesis; [1]

(ii) the number of degrees of freedom; [1]

(iii) the p-value; [2]

(iv) the conclusion, justifying your answer. [2]

(e) Using Peter's binomial model, find the probability that Argentina will win at least one of their next four international football matches. [2]

Helen thinks that a better prediction might be made by considering the transition between matches. To keep the model simple, she decides to use only two states: Argentina won (A) or Argentina did not win (B). Helen looks at Peter's list of results and counts the number of times that:

- Argentina won, twice in succession (AA),
- Argentina won, then did not win (AB),
- Argentina did not win, then won (BA),
- Argentina did not win, twice in succession (BB).

She recorded the following results.

Transition	Frequency
AA	85
AB	60
BA	62
BB	32

Helen uses the relative frequencies to estimate the probabilities in a transition matrix.

(f) (i) Given that Argentina won the previous match, show that Helen's estimate for the probability of Argentina winning the next match is $\frac{17}{29}$. [2]

(ii) Write down the transition matrix, $\boldsymbol{T}$, for Helen's model. [2]

(g) (i) Show that the characteristic polynomial of $\boldsymbol{T}$ is $1363\lambda^2 - 1263\lambda - 100 = 0$. [3]

(ii) Hence or otherwise, find the eigenvalues of $\boldsymbol{T}$. [1]

(iii) Find the corresponding eigenvectors. [3]

(h) In her retirement, many years from now, Helen is planning to travel to three consecutive international football matches involving Argentina. Use Helen's model to find the probability that Argentina will win all three matches. [4]