Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics: applications and interpretation
Higher level
Paper 3

31 October 2023

Zone A afternoon | **Zone B** afternoon | **Zone C** afternoon

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

8823 – 7203

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 28]

This question uses differential equations to model the maximum velocity of a skydiver in free fall.

In 2012, Felix Baumgartner jumped from a height of $40\,000\,\text{m}$. He was attempting to travel at the speed of sound, $330\,\text{m}\,\text{s}^{-1}$, whilst free-falling to the Earth.

Before making his attempt, Felix used mathematical models to check how realistic his attempt would be. The simplest model he used suggests that

$$\frac{\mathrm{d}v}{\mathrm{d}t} = g$$

where $v\,\text{m}\,\text{s}^{-1}$ is Felix's velocity and $g\,\text{m}\,\text{s}^{-2}$ is the acceleration due to gravity. The time since he began to free-fall is t seconds and the displacement from his initial position is s metres.

Throughout this question, the direction towards the centre of the Earth is taken to be positive and v is a positive quantity.

When $s = 0$, it is given that Felix jumps with an initial velocity $v = 10$.

(a) (i) Use the chain rule to show that $\dfrac{\mathrm{d}v}{\mathrm{d}t} = v\dfrac{\mathrm{d}v}{\mathrm{d}s}$. [1]

(ii) Assuming that g is a constant, solve the differential equation $v\dfrac{\mathrm{d}v}{\mathrm{d}s} = g$ to find v as a function of s. [4]

(iii) Using $g = 9.8$, determine whether the model predicts that Felix will succeed in travelling at the speed of sound at some point before $s = 40\,000$. Justify your answer. [3]

(b) To test the model

$$\frac{\mathrm{d}v}{\mathrm{d}t} = g\,,$$

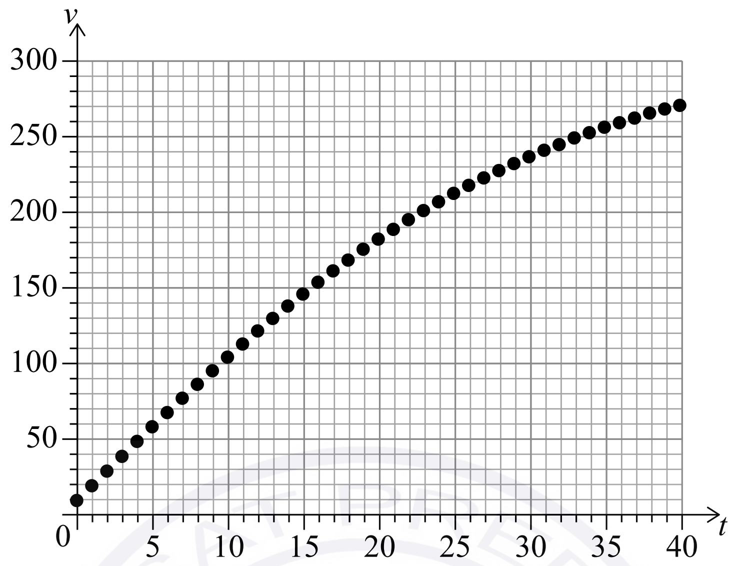
Felix conducted a trial jump from a lower height, and data for v against t was found.

(i) If the model is correct, describe the shape of the graph of v against t. [2]

(This question continues on the following page)

(Question 1 continued)

Felix's data are plotted on the following graph.



(ii) Use the plot to comment on the validity of the model in part (a). [1]

(c) An improved model considers air resistance, using

$$\frac{\mathrm{d}v}{\mathrm{d}t} = g - kv^2$$

where k is a positive constant. You are reminded that initially $s = 0$ and $v = 10$.

(i) By using $\frac{\mathrm{d}v}{\mathrm{d}t} = v\frac{\mathrm{d}v}{\mathrm{d}s}$, solve the differential equation to find v in terms of s, g and k. You may assume that $g - kv^2 > 0$. [5]

Felix uses the graph of v against t shown in part (b) to estimate the value of k.

(ii) The gradient is estimated to be 9.672 when $v = 40$. Taking g to be 9.8, use this information to show that Felix found that $k = 8 \times 10^{-5}$. [2]

(iii) Hence, find the value of v predicted by this model, as s tends to infinity. [2]

(iv) Find the upper bound for the velocity according to this model, given that $0 < s \leq 40\,000$. Give your answer to four significant figures. [2]

(This question continues on the following page)

(Question 1 continued)

The assumption that the value of g is constant is not correct. It can be shown that

$$g = \frac{3.98 \times 10^{14}}{\left(6.41 \times 10^{6} - s\right)^{2}}.$$

Hence, the new model is given by

$$v\frac{\mathrm{d}v}{\mathrm{d}s} = \frac{3.98 \times 10^{14}}{\left(6.41 \times 10^{6} - s\right)^{2}} - \left(8 \times 10^{-5}\right)v^{2}.$$

When $s = 0$, it is known that $v = 10$.

(d) Use Euler's method with a step length of 4000 to estimate the value of v when $s = 40\,000$. [4]

(e) After Felix completed his record-breaking jump, he found that the answer from part (d) was not supported by data collected during the jump.

(i) Suggest **one** improvement to the use of Euler's method which might increase the accuracy of the prediction of the model. [1]

(ii) Suggest **one** factor **not** explicitly considered by the model in part (d) which might lead to a difference between the model's prediction and the data collected. [1]

2. [Maximum mark: 27]

This question is about applying ideas from logarithms, calculus and probability to an unfamiliar mathematical theory called information theory.

Claude Shannon developed a mathematical theory called information theory to measure the information gained when random events occur. He defined the information, I, that is gained when an event with probability p occurs as

$$I = -\ln p$$

where $0 < p \leq 1$. For example, no information is gained ($I = 0$) when an event is certain to occur ($p = 1$).

(a) (i) Sketch the graph of $I = -\ln p$, for $0 < p \leq 1$, labelling all axes intercepts and asymptotes. [3]

(ii) Show, using calculus, that I is a decreasing function of p. [3]

(iii) Interpret what "I is a decreasing function of p" means in the given context. [1]

(b) A computer selects at random an integer x from 1 to 10, inclusive. Each outcome is equally likely.

Alessia is trying to determine the value of x and asks if x is odd.

(i) Write down the probability that x is odd. [1]

(ii) Alessia is told that x is odd. Find how much information Alessia gains. [2]

The computer then selects at random an integer y from 1 to 10, inclusive. Each outcome is equally likely.

Daniel is trying to determine the value of y and asks if y is 7. He is told that it is not 7.

(iii) Find how much information Daniel gains. [2]

(This question continues on the following page)

(Question 2 continued)

If a random variable has n possible outcomes with probabilities $p_1, p_2 \ldots p_n$, then the expected information gained, $\mathrm{E}(I)$, is defined as

$$\mathrm{E}(I) = \sum_{r=1}^{n} -p_r \ln p_r .$$

(c) For the integer guessing game described in part (b), when Daniel asks if y is 7, there are two possible outcomes: "y is 7" or "y is not 7".

(i) Show that the expected information gained by Daniel is 0.325, correct to three significant figures. [2]

(ii) Alessia asks if x is odd. Show that her expected information gained is greater than Daniel's expected information gained. [2]

Information theory can be applied to a variety of situations.

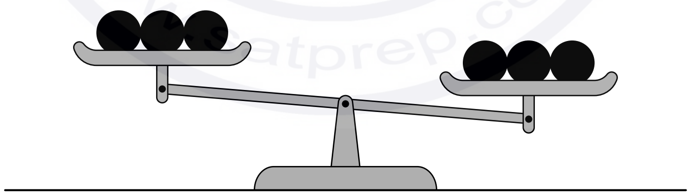
(d) When a coin is flipped, the outcome is either heads or tails. The coin may be biased. Let p be the probability of the outcome being heads.

(i) Find, in terms of p, the information gained when the outcome is tails. [1]

(ii) Find, in terms of p, the expected information gained when the coin is flipped once. [1]

(iii) Hence, find the value of p when the expected information gained is maximized. [2]

A famous puzzle uses 12 balls which appear identical. 11 have the same weight, but one is either lighter or heavier than the others. A pair of scales can be repeatedly used to compare the weights of different combinations of the balls.



The outcome of each weighing can be "balanced", "left-hand side heavier" or "right-hand side heavier". The aim of the puzzle is to identify the ball which is the different weight, and whether it is heavier or lighter than the others, in as few weighings as possible.

(This question continues on the following page)

(Question 2 continued)

(e) Angela wants to decide how many balls should be compared to each other in the **first weighing**. She produces the following table to help plan her strategy.

Number of balls on each side	**Probability of balanced**	**Probability of left-hand side heavier**	**Probability of right-hand side heavier**	**Expected information, $\mathrm{E}(I)$ (3 decimal places)**
1	$\frac{5}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	0.566
2	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\boldsymbol{z}$
3	$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{y}$	1.040
4	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1.099
5	$\frac{1}{6}$	$\frac{5}{12}$	$\frac{5}{12}$	1.028
6	0	$\frac{1}{2}$	$\frac{1}{2}$	0.693

(i) Find the value of x. Justify your answer. [2]

(ii) Find the value of y. [2]

(iii) Find the value of z. [2]

(iv) Use the table to suggest the best choice for Angela's first weighing. Justify your answer. [1]

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics: applications and interpretation
Higher level
Paper 3

9 May 2023

Zone A afternoon | **Zone B** morning | **Zone C** afternoon

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

2223 – 7213

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

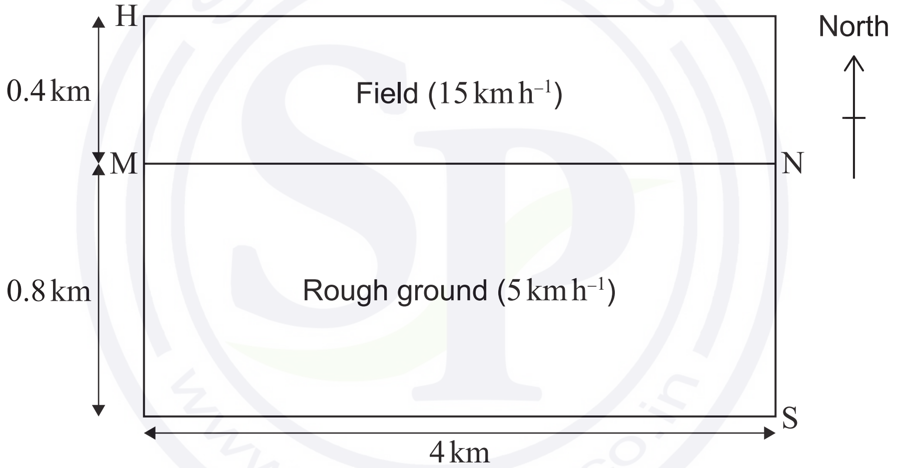
Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 26]

This question considers the optimal route between two points, separated by several regions where different speeds are possible.

Huw lives in a house, H, and he attends a school, S, where H and S are marked on the following diagram. The school is situated $1.2\,\text{km}$ south and $4\,\text{km}$ east of Huw's house. There is a boundary [MN], going from west to east, $0.4\,\text{km}$ south of his house. The land north of [MN] is a field over which Huw runs at 15 kilometres per hour $(\text{km}\,\text{h}^{-1})$. The land south of [MN] is rough ground over which Huw walks at $5\,\text{km}\,\text{h}^{-1}$. The two regions are shown in the following diagram.

diagram not to scale



(a) Huw travels in a straight line from H to S. Calculate the time that Huw takes to complete this journey. Give your answer correct to the nearest minute. [6]

(b) Huw realizes that his journey time could be reduced by taking a less direct route. He therefore defines a point P on [MN] that is $x\,\text{km}$ east of M. Huw decides to run from H to P and then walk from P to S. Let $T(x)$ represent the time, in hours, taken by Huw to complete the journey along this route.

(i) Show that $T(x) = \dfrac{\sqrt{0.4^2 + x^2} + 3\sqrt{0.8^2 + (4-x)^2}}{15}$. [3]

(ii) Sketch the graph of $y = T(x)$. [2]

(This question continues on the following page)

(Question 1 continued)

(iii) Hence determine the value of x that minimizes $T(x)$. [1]

(iv) Find by how much Huw's journey time is reduced when he takes this optimal route, compared to travelling in a straight line from H to S. Give your answer correct to the nearest minute. [2]

(c) (i) Determine an expression for the derivative $T'(x)$. [3]

(ii) Hence show that $T(x)$ is minimized when

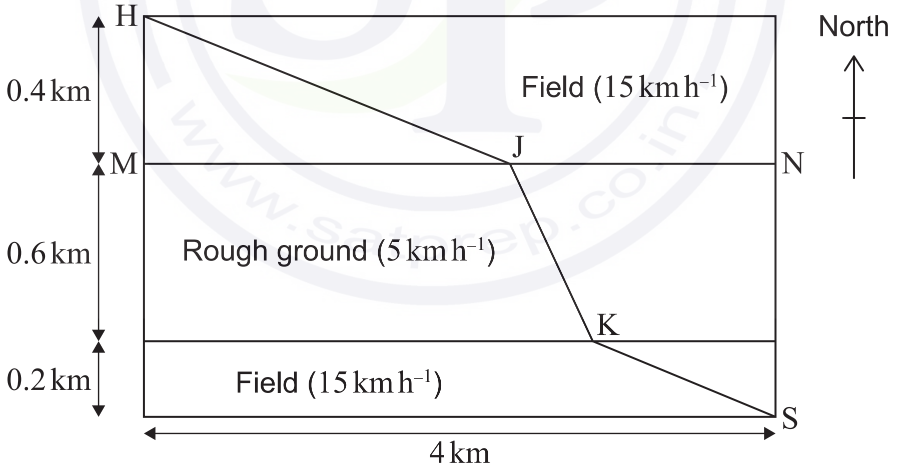
$$\frac{x}{\sqrt{0.16+x^2}} = \frac{3(4-x)}{\sqrt{0.64+(4-x)^2}}.$$ [1]

(iii) For the optimal route, verify that the equation in part (c)(ii) satisfies the following result:

$$\frac{\cos \mathrm{H\hat{P}M}}{\cos \mathrm{S\hat{P}N}} = \frac{\text{speed over field}}{\text{speed over rough ground}}.$$ [2]

(d) The owner of the rough ground converts the southern quarter into a field over which Huw can run at $15\,\mathrm{km\,h^{-1}}$. The following diagram shows the optimal route, HJKS, in this new situation. You are given that [HJ] is parallel to [KS].

diagram not to scale



Using a similar result to that given in part (c)(iii), at the point J, determine MJ. [6]

2. [Maximum mark: 29]

This question considers the analysis of several datasets of examination marks using a variety of standard procedures and also an unfamiliar statistical test.

A class of eight students sits two examinations, one in French and one in German.

The marks in these examinations are given in **Table 1**.

Table 1

Student	**French mark**	**German mark**
S_1	42	39
S_2	65	66
S_3	82	71
S_4	50	53
S_5	48	32
S_6	73	59
S_7	34	40
S_8	59	56

The maximum mark in both examinations is the same.

You may assume that these data are a random sample from a bivariate normal distribution with mean μ_{F} for the French examination, mean μ_{G} for the German examination and Pearson's product-moment correlation coefficient ρ.

Before the examinations were sat, the Head of Languages, Pierre, decided to investigate whether there would be significant evidence of a difference between μ_{F} and μ_{G}. He decided to analyse the data using a two-tailed paired t-test with significance level 5%.

(a) Explain briefly

(i) why he chose to use a t-test and not a z-test; [1]

(ii) why he chose to use a two-tailed test and not a one-tailed test. [1]

(b) (i) State suitable hypotheses for the t-test. [1]

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

(iii) The p-value is a probability. State the event for which it gives the probability. [1]

(iv) State, giving a reason, what conclusion Pierre should reach. [2]

(This question continues on the following page)

(Question 2 continued)

(c) Pierre believes that students who score well in one language examination tend to score well in the other language examination. He therefore decides to carry out a test at the 5% significance level to investigate whether there is a positive correlation between the French examination marks and the German examination marks.

(i) State appropriate hypotheses in terms of ρ. [1]

(ii) Perform a suitable test and state the p-value. State, in context, the conclusion that Pierre should reach, giving a reason. [4]

(d) There are actually two more students in this particular class, Paul and Sue. Paul sat the French examination, but he was unable to sit the German examination. Sue sat the German examination, but she was unable to sit the French examination.

(i) Paul's mark in the French examination was 58. Use the data in **Table 1** to predict the mark that Paul would have obtained in his German examination. [3]

(ii) Based on her mark in the German examination, Sue's mark in the French examination was predicted to be 71. Find the mark she obtained in the German examination. [2]

Six students sit examinations in mathematics and history and their marks are shown in **Table 2**. The Vice Principal, Angela, decides to investigate whether there is any association between the marks obtained in these two subjects.

Table 2

Student	Mathematics mark (x)	History mark (y)
P_1	53	41
P_2	76	70
P_3	50	62
P_4	65	47
P_5	61	66
P_6	84	50

(This question continues on the following page)

(Question 2 continued)

Angela is informed that the maximum mark in each subject is 100.

Angela believes that the data might not be normally distributed, so she investigates what suitable tests are available which do not assume the data is normally distributed. She decides to use an unfamiliar test based on a statistic called Kendall's τ.

Consider n bivariate observations (x_i, y_i), $i = 1, 2, \ldots n$, such that there are no equal x-values and no equal y-values. Any pair of distinct bivariate observations (x_i, y_i) and (x_j, y_j) is said to be concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$. For n bivariate observations, there are $\frac{n(n-1)}{2}$ distinct pairs.

Kendall's τ is defined as $\frac{2(C-D)}{n(n-1)}$ where C and D denote respectively the number of concordant pairs and discordant pairs.

(e) (i) Show that the value of Kendall's τ always lies in the interval $[-1, +1]$. [1]

(ii) For students P_1 and P_2, show that their pair is concordant. [1]

(iii) Show that the value of Kendall's τ for the mathematics and history data is 0.2. [4]

(f) Angela decided to use this statistic in a two-tailed test at the 10% significance level. The critical region for her test is $|\tau| \geq 0.733$.

(i) State, in words, her null and alternative hypotheses. [1]

(ii) State the conclusion that Angela should reach. Give a reason for your answer. [2]

Angela now finds that the history marks are actually out of 120. The history teacher advises Angela to scale the history marks so that they are out of 100 and then redo the calculations for the value of τ.

(g) State, with a reason, whether you agree with the advice given by the history teacher. [2]

References:

Mathematics: applications and interpretation
Higher level
Paper 3

9 May 2023

Zone A afternoon | **Zone B** morning | **Zone C** afternoon

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

2223 – 7208

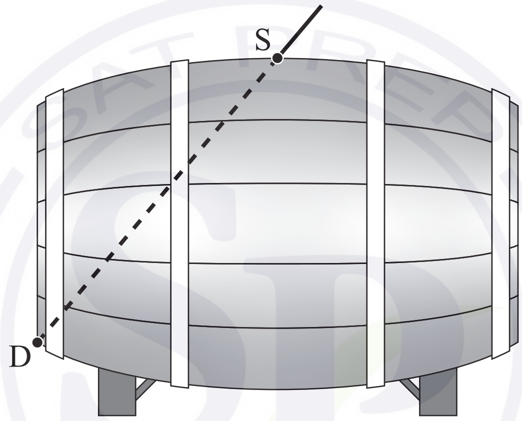
Blank page

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 26]

In this question you will use a historic method of calculating the cost of a barrel of wine to determine which shape of barrel gives the best value for money.

In Austria in the 17th century, one method for measuring the volume of a barrel of wine, and hence determining its cost, was by inserting a straight stick into a hole in the side, as shown in the following diagram, and measuring the length SD. The longer the length, the greater the cost to the customer.



Let SD be d metres and the cost be C gulden (the local currency at the time). When the length of SD was 0.5 metres, the cost was 0.80 gulden.

(a) Given that C was directly proportional to d, find an equation for C in terms of d. [3]

A particular barrel of wine cost 0.96 gulden.

(b) Show that $d = 0.6$. [1]

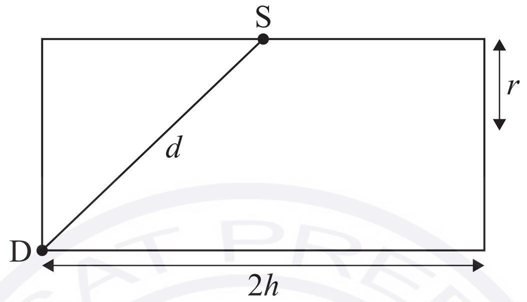
(This question continues on the following page)

(Question 1 continued)

This method of determining the cost was noticed by a mathematician, Kepler, who decided to try to calculate the dimensions of a barrel which would give the maximum volume of wine for a given length SD.

Initially he modelled the barrel as a cylinder, with S at the midpoint of one side. He took the length of the cylinder as $2h$ metres and its radius as r metres, as shown in the following diagram of the cross-section.

diagram not to scale



(c) Find an expression for r^2 in terms of d and h. [3]

Let the volume of this barrel be $V\,\mathrm{m}^3$.

(d) Show that $V = \frac{\pi}{2}\left(d^2h - h^3\right)$. [2]

The remainder of this question considers the shape of barrel that gives the best value when $d = 0.6$.

(e) (i) Use the formula from part (d) to find the volume of this barrel when $h = 0.4$. [2]

(ii) Use differentiation to show that $h = \sqrt{0.12}$ when $\frac{\mathrm{d}V}{\mathrm{d}h} = 0$. [3]

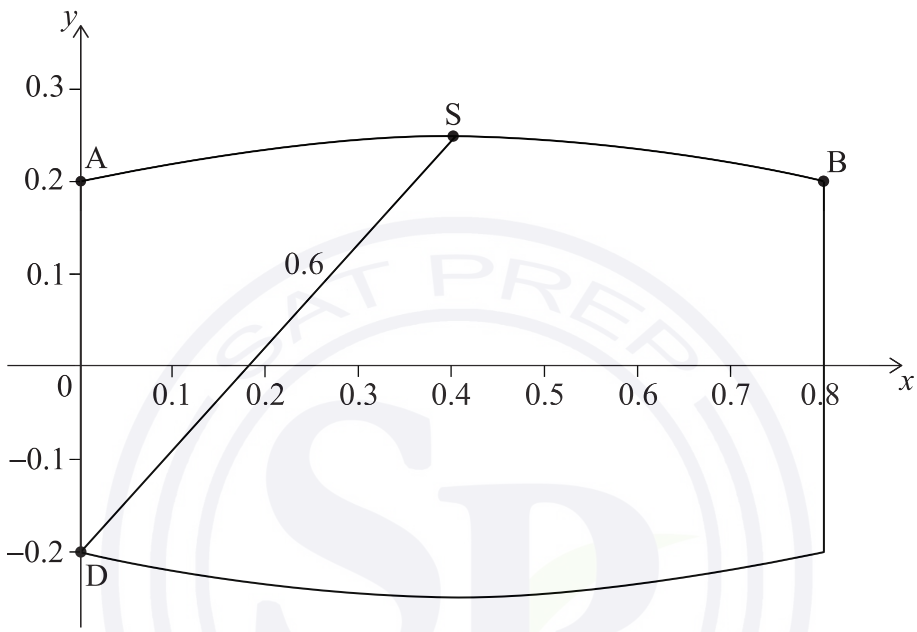
(iii) Given that this value of h maximizes the volume, find the largest possible volume of this barrel. [2]

(This question continues on the following page)

(Question 1 continued)

Kepler then considered a non-cylindrical barrel whose base and lid are circles with radius $0.2\,\text{m}$ and whose length is $0.8\,\text{m}$.

He modelled the curved surface of this barrel by rotating a quadratic curve, ASB, with equation $y = ax^2 + bx + c$, $0 \le x \le 0.8$, about the x-axis. The origin of the coordinate system is at the centre of one of the circular faces as shown in the following diagram. S is at the vertex of the quadratic curve and $\text{SD} = 0.6$.



Kepler wished to find out if his barrel would give him more wine than any cylindrical barrel with $d = 0.6$.

The coordinates of A and B are $(0,\ 0.2)$ and $(0.8,\ 0.2)$ respectively.

(f) Find the equation of the quadratic curve, ASB. [6]

(g) Show that the volume of this barrel is greater than the maximum volume of any cylindrical barrel with $d = 0.6$. [3]

(h) State one assumption, not already given, that has been made in using these models to find the shape of the barrel that gives the best value. [1]

2. [Maximum mark: 29]

In this question you will use vector methods to determine whether aircraft are obeying air traffic regulations.

The base of an air traffic control tower at an airport is taken as the origin of a coordinate system. An aircraft's position is given by the coordinates (x, y, z), where x and y are respectively the aircraft's displacement east and north of the tower, and z is the vertical displacement of the aircraft above the base of the tower. All displacements are measured in kilometres.

At $12{:}00$ two aircraft, A and B, are at the points $\mathrm{P}(100, -82, 10.7)$ and $\mathrm{Q}(215, -197, 10.7)$ respectively.

(a) Find the distance between the two aircraft at $12{:}00$. [2]

The two aircraft are flying along the same straight line (flight path), with B behind A.

They both have the same constant velocity of $\begin{pmatrix} -640 \\ 640 \\ 0 \end{pmatrix}$ kilometres per hour.

(b) Find the speed of both aircraft. [2]

Air traffic regulations state that if two aircraft are on the same flight path then they must always maintain at least a 10 minute gap between them. If at any time two aircraft are too close they are said to be "in conflict".

(c) Find the length of time it takes B to reach point P from point Q, and hence state whether the two aircraft are in conflict. [3]

(d) Write down, $\boldsymbol{r}_A$, the position vector of A, t hours after $12{:}00$. [1]

If two aircraft are not on the same flight path, air traffic regulations state:

When the vertical distance between the two aircraft is less than $300\,\mathrm{m}$ the aircraft must be more than $10\,\mathrm{km}$ apart.

When the vertical distance between the two aircraft is at least $300\,\mathrm{m}$ there are no restrictions.

The air traffic controller notices an aircraft, C, flying on a different flight path but close to A. The position of C, t hours after $12{:}00$, is given by

$$\boldsymbol{r}_C = \begin{pmatrix} -400 \\ -41 \\ 9.1 \end{pmatrix} + t\begin{pmatrix} -140 \\ 604 \\ 2 \end{pmatrix}.$$

(This question continues on the following page)

(Question 2 continued)

(e) (i) Find the two values of t at which the distance between A and C is $10\,\text{km}$. [5]

It is given that the distance between A and C is less than $10\,\text{km}$, only between these two values of t.

(ii) Determine whether the two aircraft, A and C, will break the air traffic regulations if they continue with their current velocities. Justify your answer. [5]

A new coordinate system, (x, y), is defined with an origin R at a point $2\,\text{km}$ directly above the air traffic control tower. When an aircraft is flying in the horizontal plane containing R, the values of x and y represent its displacement, in km, east and north of point R respectively.

A fourth aircraft, D, is flying at a constant height of $2\,\text{km}$ near the airport while waiting for permission to land. Its position at time t is given by

$$\overrightarrow{\text{RD}} = \begin{pmatrix} 6.4\cos(38t) \\ 6.4\sin(38t) \end{pmatrix}.$$

(f) Describe the path followed by D. [3]

Around the same time a small aircraft, E, is flying at the same height as D and along the line with vector equation

$$\overrightarrow{\text{RE}} = \begin{pmatrix} 20 \\ 10 \end{pmatrix} + \lambda \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

(g) Let $\boldsymbol{b} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

(i) Find $\overrightarrow{\text{RE}} \bullet \boldsymbol{b}$ in terms of λ. [2]

(ii) Hence find the value of λ for which the distance from R to the line is minimum. [2]

(iii) Find this minimum distance. [2]

(iv) Show that D and E will not break the air traffic regulations for any value of t. [2]

References:

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics: applications and interpretation
Higher level
Paper 3

Tuesday 8 November 2022 (afternoon)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

8 pages

8822 – 7203

Blank page

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [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]

(This question continues on the following page)

(Question 1 continued)

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]

(This question continues on the following page)

(Question 1 continued)

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]

2. [Maximum mark: 26]

Some medical conditions require patients to take medication regularly for long periods of time. In this question, you will explore the concentration of a medicinal drug in the body, when the drug is given repeatedly.

Once a drug enters the body, it is absorbed into the blood. As the body breaks down the drug over time, the concentration of the drug decreases. Let $C(t)$, measured in milligrams per millilitre $(\mathrm{mg\,ml^{-1}})$, be the concentration of the drug, t hours after the drug is given to the patient. The rate at which the drug is broken down is modelled as directly proportional to its concentration, leading to the differential equation

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -kC \text{, where } k \in \mathbb{R}^{+}.$$

The initial concentration is $d\ \mathrm{mg\,ml^{-1}}$, $d > 0$.

(a) By solving the differential equation, show that $C = d\mathrm{e}^{-kt}$. [3]

For the remainder of this question, you will consider a particular drug where it is known that $k = 0.2$. The first dose is given at time $t = 0$ and it is assumed that before this there is no drug present in the blood.

(b) Find the time, in hours, for this drug to reach 5% of its initial concentration. [2]

The drug is to be given every T hours and in constant doses, such that the concentration of the drug is increased by an amount $d\ \mathrm{mg\,ml^{-1}}$. To simplify the model, it is assumed that each time the drug is given the concentration of the drug in the blood increases instantaneously.

(c) Show that the concentration of the drug is $d\left(1+\mathrm{e}^{-0.2T}+\mathrm{e}^{-0.4T}\right)$ immediately after the third dose is given. [4]

Immediately after the nth dose is given, the concentration of the drug is

$$d\left(1+\mathrm{e}^{-0.2T}+\mathrm{e}^{-0.4T}+\ldots+\mathrm{e}^{-0.2(n-1)T}\right).$$

(d) Show that this concentration can be expressed as $d\left(\dfrac{1-\mathrm{e}^{-0.2nT}}{1-\mathrm{e}^{-0.2T}}\right)$. [2]

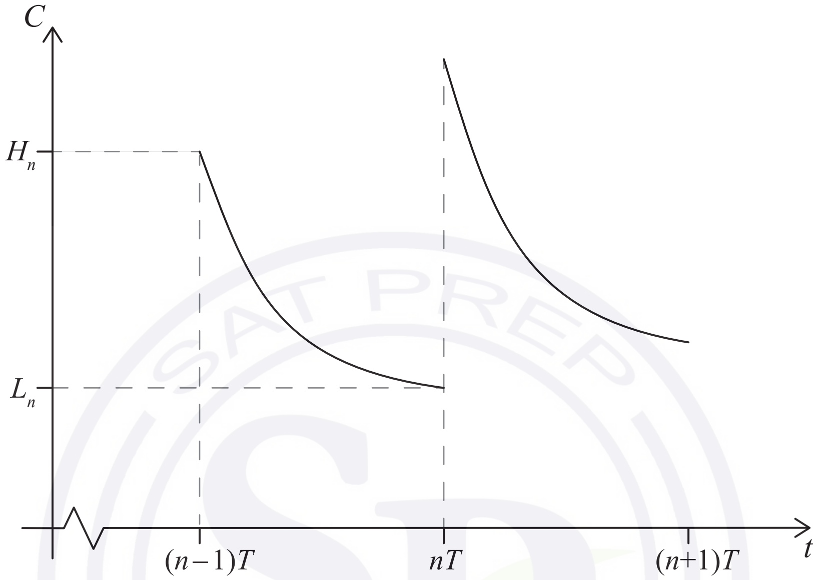
(This question continues on the following page)

(Question 2 continued)

After a patient has been taking this drug for a long time, it is required to keep the concentration within a particular range so that it is both safe and effective.

Let H_n be the highest concentration of the drug in the body for the interval $(n-1)T \le t < nT$.

Let L_n be the lowest concentration of the drug in the body for the interval $(n-1)T \le t < nT$. This is shown in the following graph.



H_∞ is defined as $\lim\limits_{n\to\infty} H_n$ and L_∞ is defined as $\lim\limits_{n\to\infty} L_n$.

(e) Find, in terms of d and T, an expression for

(i) H_∞. [2]

(ii) L_∞. [3]

(f) Show that

(i) $H_\infty - L_\infty = d$. [2]

(ii) $5\ln\left(\dfrac{H_\infty}{L_\infty}\right) = T$. [3]

(This question continues on the following page)

(Question 2 continued)

It is known that this drug is ineffective if the long-term concentration is less than $0.06\,\text{mg}\,\text{ml}^{-1}$ and safe if it never exceeds $0.28\,\text{mg}\,\text{ml}^{-1}$.

(g) Hence, for this drug, find a suitable value for

(i) d. [1]

(ii) T. [1]

(h) For the values of d and T found in part (g), find the proportion of time for which the concentration of the drug is at least $0.06\,\text{mg}\,\text{ml}^{-1}$ between the first and second doses. [2]

(i) Suggest a reason why the instructions on the label of the drug might use a different value for T to that found in part (g)(ii). [1]

References:

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics: applications and interpretation
Higher level
Paper 3

Thursday 12 May 2022 (morning)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

7 pages

2222–7213

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [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]

(This question continues on the following page)

(Question 1 continued)

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]

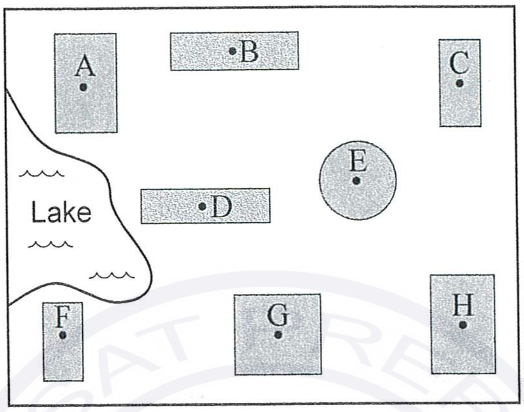
(f) By considering the claims of both Aimmika and Nichakarn, explain whether the advertising was beneficial to the store. [3]

Blank page

2. [Maximum mark: 28]

This question compares possible designs for a new computer network between multiple school buildings, and whether they meet specific requirements.

A school's administration team decides to install new fibre-optic internet cables underground. The school has eight buildings that need to be connected by these cables. A map of the school is shown below, with the internet access point of each building labelled A–H.



Jonas is planning where to install the underground cables. He begins by determining the distances, in metres, between the underground access points in each of the buildings.

He finds $\mathrm{AD} = 89.2\,\mathrm{m}$, $\mathrm{DF} = 104.9\,\mathrm{m}$ and $\mathrm{A\hat{D}F} = 83^\circ$.

(a) Find AF. [3]

The cost for installing the cable directly between A and F is $\$21\,310$.

(b) Find the cost per metre of installing this cable. [2]

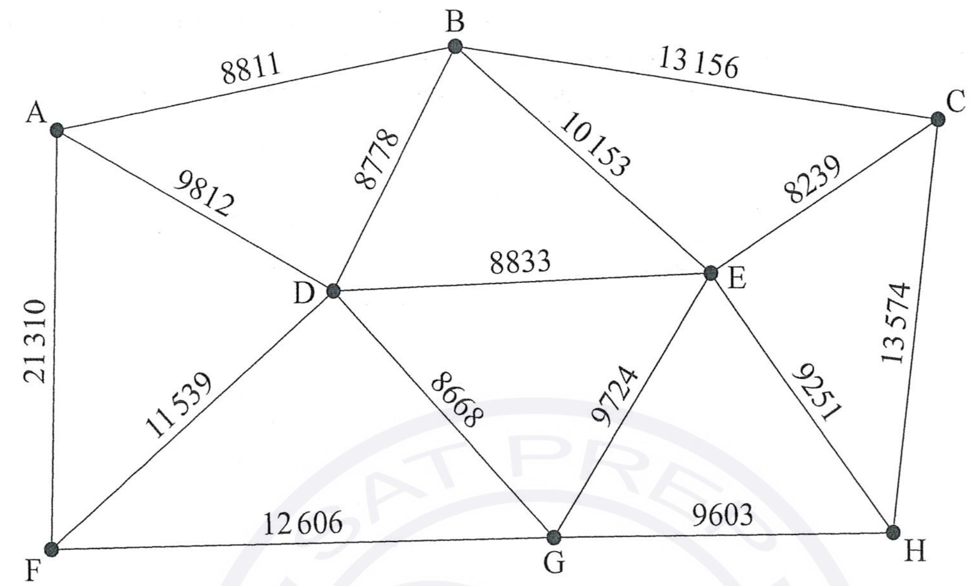
Jonas estimates that it will cost $\$110$ per metre to install the cables between all the other buildings.

(c) State why the cost for installing the cable between A and F would be higher than between the other buildings. [1]

(This question continues on the following page)

(Question 2 continued)

Jonas creates the following graph, S, using the cost of installing the cables between two buildings as the weight of each edge.



The computer network could be designed such that each building is directly connected to at least one other building and hence all buildings are indirectly connected.

(d) (i) By using Kruskal's algorithm, find the minimum spanning tree for S, showing clearly the order in which edges are added. [3]

(ii) Hence find the minimum installation cost for the cables that would allow all the buildings to be part of the computer network. [2]

The computer network fails if any part of it becomes unreachable from any other part. To help protect the network from failing, every building could be connected to at least two other buildings. In this way if one connection breaks, the building is still part of the computer network. Jonas can achieve this by finding a Hamiltonian cycle within the graph.

(e) State why a path that forms a Hamiltonian cycle does not always form an Eulerian circuit. [1]

(f) Starting at D, use the nearest neighbour algorithm to find the upper bound for the installation cost of a computer network in the form of a Hamiltonian cycle.

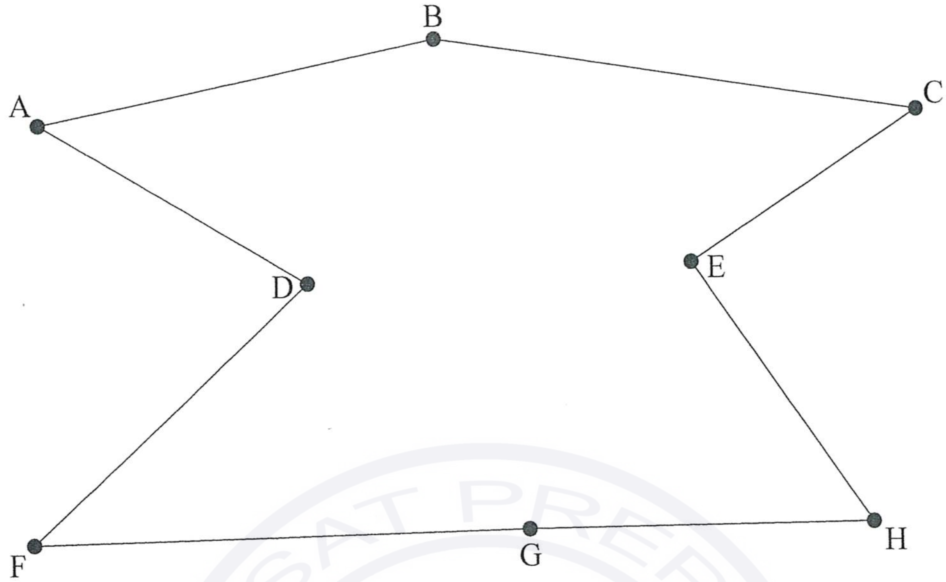
Note: Although the graph is not complete, in this instance it is not necessary to form a table of least distances. [5]

(g) By deleting D, use the deleted vertex algorithm to find the lower bound for the installation cost of the cycle. [6]

(This question continues on the following page)

(Question 2 continued)

After more research, Jonas decides to install the cables as shown in the diagram below.



Each individual cable is installed such that each end of the cable is connected to a building's access point. The connection between each end of a cable and an access point has a 1.4% probability of failing after a power surge.

For the network to be successful, each building in the network must be able to communicate with every other building in the network. In other words, there must be a path that connects any two buildings in the network. Jonas would like the network to have less than a 2% probability of failing to operate after a power surge.

(h) Show that Jonas's network satisfies the requirement of there being less than a 2% probability of the network failing after a power surge. [5]

Mathematics: applications and interpretation
Higher level
Paper 3

Thursday 12 May 2022 (morning)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

2222 – 7208

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 28]

This question is about modelling the spread of a computer virus to predict the number of computers in a city which will be infected by the virus.

A systems analyst defines the following variables in a model:

- t is the number of days since the first computer was infected by the virus.
- $Q(t)$ is the total number of computers that have been infected up to and including day t.

The following data were collected:

t	10	15	20	25	30	35	40
$Q(t)$	20	90	403	1806	8070	32 667	120 146

(a) (i) Find the equation of the regression line of $Q(t)$ on t. [2]

(ii) Write down the value of r, Pearson's product-moment correlation coefficient. [1]

(iii) Explain why it would not be appropriate to conduct a hypothesis test on the value of r found in (a)(ii). [1]

A model for the early stage of the spread of the computer virus suggests that

$$Q'(t) = \beta N Q(t)$$

where N is the total number of computers in a city and β is a measure of how easily the virus is spreading between computers. Both N and β are assumed to be constant.

(b) (i) Find the general solution of the differential equation $Q'(t) = \beta N Q(t)$. [4]

(ii) Using the data in the table write down the equation for an appropriate non-linear regression model. [2]

(iii) Write down the value of R^2 for this model. [1]

(iv) Hence comment on the suitability of the model from (b)(ii) in comparison with the linear model found in part (a). [2]

(v) By considering large values of t write down one criticism of the model found in (b)(ii). [1]

(This question continues on the following page)

(Question 1 continued)

(c) Use your answer from part (b)(ii) to estimate the time taken for the number of infected computers to double. [2]

The data above are taken from city X which is estimated to have 2.6 million computers. The analyst looks at data for another city, Y. These data indicate a value of $\beta = 9.64 \times 10^{-8}$.

(d) Find in which city, X or Y, the computer virus is spreading more easily. Justify your answer using your results from part (b). [3]

An estimate for $Q'(t)$, $t \geq 5$, can be found by using the formula:

$$Q'(t) \approx \frac{Q(t+5) - Q(t-5)}{10}.$$

The following table shows estimates of $Q'(t)$ for city X at different values of t.

t	10	15	20	25	30	35	40
$Q(t)$	20	90	403	1806	8070	32 667	120 146
$Q'(t)$		a	171.6	766.7	b	11 207.6	

(e) Determine the value of a and of b. Give your answers correct to one decimal place. [2]

An improved model for $Q(t)$, which is valid for large values of t, is the logistic differential equation

$$Q'(t) = kQ(t)\left(1 - \frac{Q(t)}{L}\right)$$

where k and L are constants.

Based on this differential equation, the graph of $\frac{Q'(t)}{Q(t)}$ against $Q(t)$ is predicted to be a straight line.

(f) (i) Use linear regression to estimate the value of k and of L. [5]

(ii) The solution to the differential equation is given by

$$Q(t) = \frac{L}{1 + Ce^{-kt}}$$

where C is a constant.

Using your answer to part (f)(i), estimate the percentage of computers in city X that are expected to have been infected by the virus over a long period of time. [2]

2. [Maximum mark: 27]

This question is about a metropolitan area council planning a new town and the location of a new toxic waste dump.

A metropolitan area in a country is modelled as a square. The area has four towns, located at the corners of the square. All units are in kilometres with the x-coordinate representing the distance east and the y-coordinate representing the distance north from the origin at $(0, 0)$.

- Edison is modelled as being positioned at $\mathrm{E}(0, 40)$.
- Fermitown is modelled as being positioned at $\mathrm{F}(40, 40)$.
- Gaussville is modelled as being positioned at $\mathrm{G}(40, 0)$.
- Hamilton is modelled as being positioned at $\mathrm{H}(0, 0)$.

(a) The model assumes that each town is positioned at a single point. Describe possible circumstances in which this modelling assumption is reasonable. [1]

(b) Sketch a Voronoi diagram showing the regions within the metropolitan area that are closest to each town. [1]

The metropolitan area council decides to build a new town called Isaacopolis located at $\mathrm{I}(30, 20)$.

A new Voronoi diagram is to be created to include Isaacopolis. The equation of the perpendicular bisector of $[\mathrm{IE}]$ is $y = \frac{3}{2}x + \frac{15}{2}$.

(c) (i) Find the equation of the perpendicular bisector of $[\mathrm{IF}]$. [4]

(ii) Given that the coordinates of one vertex of the new Voronoi diagram are $(20, 37.5)$, find the coordinates of the other two vertices within the metropolitan area. [4]

(iii) Sketch this new Voronoi diagram showing the regions within the metropolitan area which are closest to each town. [2]

The metropolitan area is divided into districts based on the Voronoi regions found in part (c).

(d) A car departs from a point due north of Hamilton. It travels due east at constant speed to a destination point due North of Gaussville. It passes through the Edison, Isaacopolis and Fermitown districts. The car spends 30% of the travel time in the Isaacopolis district.

Find the distance between Gaussville and the car's destination point. [4]

(This question continues on the following page)

(Question 2 continued)

A toxic waste dump needs to be located within the metropolitan area. The council wants to locate it as far as possible from the nearest town.

(e) (i) Find the location of the toxic waste dump, given that this location is not on the edge of the metropolitan area. [4]

(ii) Make one possible criticism of the council's choice of location. [1]

(f) The toxic waste dump, T, is connected to the towns via a system of sewers.

The connections are represented in the following matrix, $\boldsymbol{M}$, where the order of rows and columns is (E, F, G, H, I, T).

$$\boldsymbol{M}=\begin{pmatrix}1&0&1&1&0&0\\0&1&0&0&0&1\\1&0&1&0&1&0\\1&0&0&1&0&1\\0&0&1&0&1&0\\0&1&0&1&0&1\end{pmatrix}$$

A leak occurs from the toxic waste dump and travels through the sewers. The pollution takes one day to travel between locations that are directly connected.

The digit 1 in $\boldsymbol{M}$ represents a direct connection. The values of 1 in the leading diagonal of $\boldsymbol{M}$ mean that once a location is polluted it will stay polluted.

(i) Find which town is last to be polluted. Justify your answer. [3]

(ii) Write down the number of days it takes for the pollution to reach the last town. [1]

(iii) A sewer inspector needs to plan the shortest possible route through each of the connections between different locations. Determine an appropriate start point and an appropriate end point of the inspection route.

Note that the fact that each location is connected to itself does not correspond to a sewer that needs to be inspected. [2]

References:

Mathematics: applications and interpretation
Higher level
Paper 3

Tuesday 9 November 2021 (morning)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

8821 – 7203

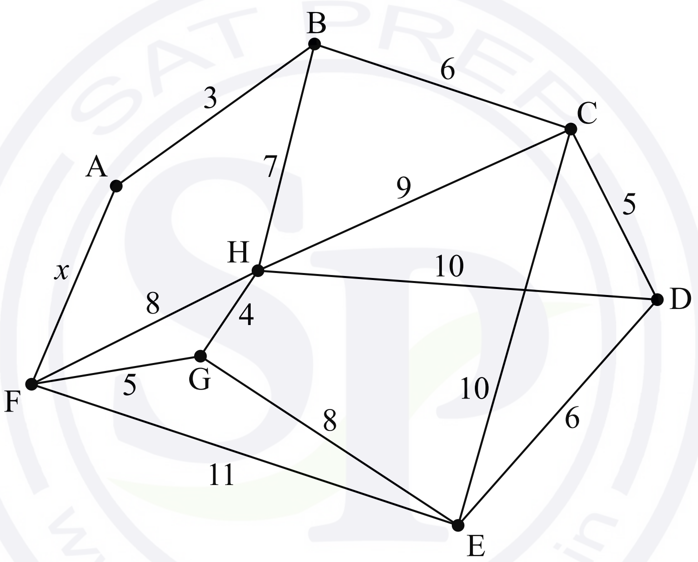
6 pages

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 25]

This question explores how graph algorithms can be applied to a graph with an unknown edge weight.

Graph W is shown in the following diagram. The vertices of W represent tourist attractions in a city. The weight of each edge represents the travel time, to the nearest minute, between two attractions. The route between A and F is currently being resurfaced and this has led to a variable travel time. For this reason, AF has an unknown travel time x minutes, where $x \in \mathbb{Z}^+$.



(a) Write down a Hamiltonian cycle in W. [1]

Daniel plans to visit all the attractions, starting and finishing at A. He wants to minimize his travel time.

To find a lower bound for Daniel's travel time, vertex A and its adjacent edges are first deleted.

(b) (i) Use Prim's algorithm, starting at vertex B, to find the weight of the minimum spanning tree of the remaining graph. You should indicate clearly the order in which the algorithm selects each edge. [5]

(ii) Hence, for the case where $x < 9$, find a lower bound for Daniel's travel time, in terms of x. [2]

(This question continues on the following page)

(Question 1 continued)

Daniel makes a table to show the minimum travel time between each pair of attractions.

	A	B	C	D	E	F	G	H
A		3	9	14	19 **or** $(11+x)$	18 **or** x	14 **or** $(5+x)$	10 **or** $(8+x)$
B			6	11	16 **or** $(14+x)$	15 **or** $(3+x)$	11 **or** $(8+x)$	7
C				5	10	17 **or** $(9+x)$	p	9
D					6	q **or** $(r+x)$	14	10
E						11	8	12
F							5	8
G								4
H								

(c) Write down the value of

(i) p; [1]

(ii) q; [1]

(iii) r. [1]

To find an upper bound for Daniel's travel time, the nearest neighbour algorithm is used, starting at vertex A.

(d) Consider the case where $x = 3$.

(i) Use the nearest neighbour algorithm to find two possible cycles. [3]

(ii) Find the best upper bound for Daniel's travel time. [2]

(e) Consider the case where $x > 3$.

(i) Find the least value of x for which the edge AF will definitely not be used by Daniel. [2]

(ii) Hence state the value of the upper bound for Daniel's travel time for the value of x found in part (e)(i). [2]

The tourist office in the city has received complaints about the lack of cleanliness of some routes between the attractions. Corinne, the office manager, decides to inspect all the routes between all the attractions, starting and finishing at H. The sum of the weights of all the edges in graph W is $(92 + x)$.

Corinne inspects all the routes as quickly as possible and takes 2 hours.

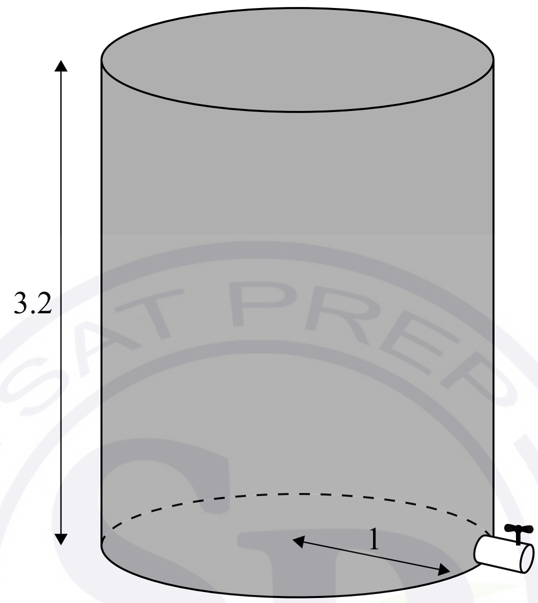
(f) Find the value of x during Corinne's inspection. [5]

2. [Maximum mark: 30]

This question explores models for the height of water in a cylindrical container as water drains out.

The diagram shows a cylindrical water container of height 3.2 metres and base radius 1 metre. At the base of the container is a small circular valve, which enables water to drain out.

diagram not to scale



Eva closes the valve and fills the container with water.

At time $t = 0$, Eva opens the valve. She records the height, h metres, of water remaining in the container every 5 minutes.

Time, t (minutes)	Height, h (metres)
0	3.2
5	2.4
10	1.6
15	1.1
20	0.5

Eva first tries to model the height using a linear function, $h(t) = at + b$, where a, $b \in \mathbb{R}$.

(a) (i) Find the equation of the regression line of h on t. [2]

(ii) Interpret the meaning of parameter a in the context of the model. [1]

(This question continues on the following page)

(Question 2 continued)

Eva uses the equation of the regression line of h on t, to predict the time it will take for all the water to drain out of the container.

(iii) Suggest why Eva's use of the linear regression equation in this way could be unreliable. [1]

Eva thinks she can improve her model by using a quadratic function, $h(t) = pt^2 + qt + r$, where p, q, $r \in \mathbb{R}$.

(b) (i) Find the equation of the least squares quadratic regression curve. [1]

Eva uses this equation to predict the time it will take for all the water to drain out of the container and obtains an answer of k minutes.

(ii) Find the value of k. [2]

(iii) Hence, write down a suitable domain for Eva's function $h(t) = pt^2 + qt + r$. [1]

Let V be the volume, in cubic metres, of water in the container at time t minutes.
Let R be the radius, in metres, of the circular valve.

Eva does some research and discovers a formula for the rate of change of V.

$$\frac{dV}{dt} = -\pi R^2 \sqrt{70\,560\,h}$$

(c) Show that $\frac{dh}{dt} = -R^2\sqrt{70\,560\,h}$. [3]

(d) By solving the differential equation $\frac{dh}{dt} = -R^2\sqrt{70\,560\,h}$, show that the general solution is given by $h = 17\,640\left(c - R^2 t\right)^2$, where $c \in \mathbb{R}$. [5]

Eva measures the radius of the valve to be 0.023 metres. Let T be the time, in minutes, it takes for all the water to drain out of the container.

(e) Use the general solution from part (d) and the initial condition $h(0) = 3.2$ to predict the value of T. [4]

Eva wants to use the container as a timer. She adjusts the initial height of water in the container so that all the water will drain out of the container in 15 minutes.

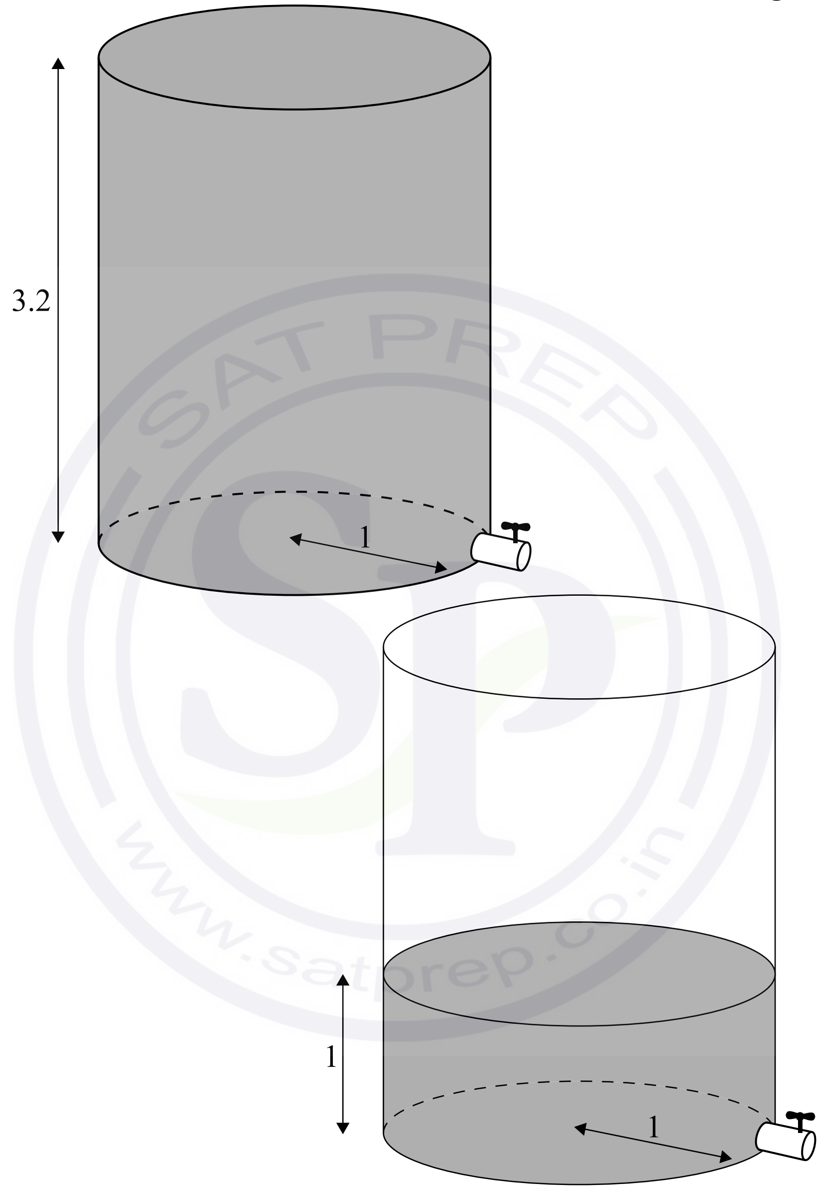
(f) Find this new height. [3]

(This question continues on the following page)

(Question 2 continued)

Eva has another water container that is identical to the first one. She places one water container above the other one, so that all the water from the highest container will drain into the lowest container. Eva completely fills the highest container, but only fills the lowest container to a height of 1 metre, as shown in the diagram.

diagram not to scale



At time $t = 0$ Eva opens both valves. Let H be the height of water, in metres, in the lowest container at time t.

(g) (i) Show that $\dfrac{\mathrm{d}H}{\mathrm{d}t} \approx 0.2514 - 0.009873t - 0.1405\sqrt{H}$, where $0 \le t \le T$. [4]

(ii) Use Euler's method with a step length of 0.5 minutes to estimate the maximum value of H. [3]

References:

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics: applications and interpretation
Higher level
Paper 3

Tuesday 11 May 2021 (morning)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [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]

(This question continues on the following page)

(Question 1 continued)

(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]

(This question continues on page 5)

Blank page

(Question 1 continued)

(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]

2. [Maximum mark: 31]

Alessia is an ecologist working for Mediterranean fishing authorities. She is interested in whether the mackerel population density is likely to fall below 5000 mackerel per km^3, as this is the minimum value required for sustainable fishing. She believes that the primary factor affecting the mackerel population is the interaction of mackerel with sharks, their main predator.

The population densities of mackerel (M thousands per km^3) and sharks (S per km^3) in the Mediterranean Sea are modelled by the coupled differential equations:

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \alpha M - \beta MS$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \gamma MS - \delta S$$

where t is measured in years, and α, β, γ and δ are parameters.

This model assumes that no other factors affect the mackerel or shark population densities.

The term αM models the population growth rate of the mackerel in the absence of sharks.
The term βMS models the death rate of the mackerel due to being eaten by sharks.

(a) Suggest similar interpretations for the following terms.

(i) γMS [1]

(ii) δS [1]

(b) An equilibrium point is a set of values of M and S, such that $\frac{\mathrm{d}M}{\mathrm{d}t} = 0$ and $\frac{\mathrm{d}S}{\mathrm{d}t} = 0$.

Given that both species are present at the equilibrium point,

(i) show that, at the equilibrium point, the value of the mackerel population density is $\frac{\delta}{\gamma}$; [3]

(ii) find the value of the shark population density at the equilibrium point. [2]

(c) The equilibrium point found in part (b) gives the average values of M and S over time.

Use the model to predict how the following events would affect the average value of M. Justify your answers.

(i) Toxic sewage is added to the Mediterranean Sea. Alessia claims this reduces the shark population growth rate and hence the value of γ is halved. No other parameter changes. [2]

(ii) Global warming increases the temperature of the Mediterranean Sea. Alessia claims that this promotes the mackerel population growth rate and hence the value of α is doubled. No other parameter changes. [2]

(This question continues on the following page)

(Question 2 continued)

(d) To estimate the value of α, Alessia considers a situation where there are no sharks and the initial mackerel population density is M_0.

(i) Write down the differential equation for M that models this situation. [1]

(ii) Show that the expression for the mackerel population density after t years is $M = M_0 e^{\alpha t}$. [4]

(iii) Alessia estimates that the mackerel population density increases by a factor of three every two years. Show that $\alpha = 0.549$ to three significant figures. [3]

Based on additional observations, it is believed that

$$\alpha = 0.549,$$

$$\beta = 0.236,$$

$$\gamma = 0.244,$$

$$\delta = 1.39.$$

Alessia decides to use Euler's method to estimate future mackerel and shark population densities. The initial population densities are estimated to be $M_0 = 5.7$ and $S_0 = 2$. She uses a step length of 0.1 years.

(e) (i) Write down expressions for M_{n+1} and S_{n+1} in terms of M_n and S_n. [3]

(ii) Use Euler's method to find an estimate for the mackerel population density after one year. [2]

(f) Alessia will use her model to estimate whether the mackerel population density is likely to fall below the minimum value required for sustainable fishing, 5000 per km^3, during the first nine years.

(i) Use Euler's method to sketch the trajectory of the phase portrait, for $4 \le M \le 7$ and $1.5 \le S \le 3$, over the first nine years. [3]

(ii) Using your phase portrait, or otherwise, determine whether the mackerel population density would be sufficient to support sustainable fishing during the first nine years. [2]

(iii) State **two** reasons why Alessia's conclusion, found in part (f)(ii), might not be valid. [2]

References:

Mathematics: applications and interpretation
Higher level
Paper 3

Tuesday 11 May 2021 (morning)

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

2221 – 7208

6 pages

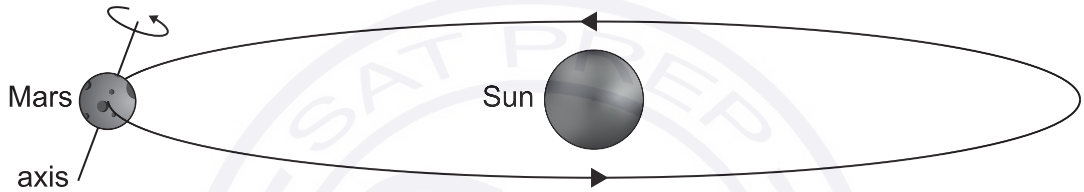
Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

1. [Maximum mark: 27]

A suitable site for the landing of a spacecraft on the planet Mars is identified at a point, A. The shortest time from sunrise to sunset at point A must be found.

Radians should be used throughout this question. All values given in the question should be treated as exact.

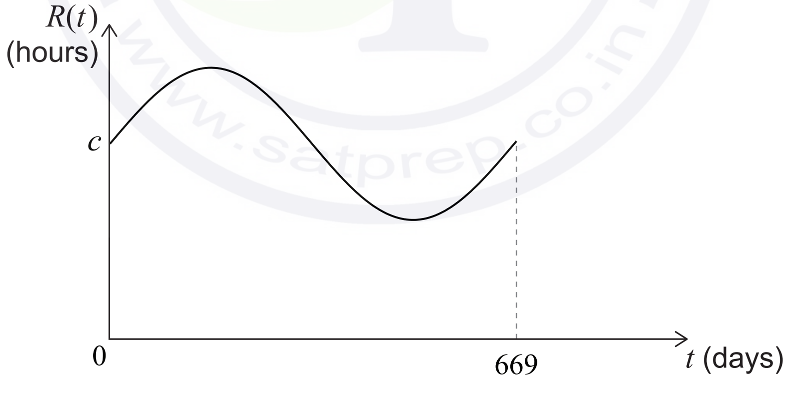
Mars completes a full orbit of the Sun in 669 Martian days, which is one Martian year.



On day t, where $t \in \mathbb{Z}$, the length of time, in hours, from the start of the Martian day until sunrise at point A can be modelled by a function, $R(t)$, where

$$R(t) = a\sin(bt) + c,\ t \in \mathbb{R}.$$

The graph of R is shown for one Martian year.



(a) Show that $b \approx 0.00939$. [2]

Mars completes a full rotation on its axis in 24 hours and 40 minutes.

(b) Find the angle through which Mars rotates on its axis each hour. [3]

(This question continues on the following page)

(Question 1 continued)

The time of sunrise on Mars depends on the angle, δ, at which it tilts towards the Sun. During a Martian year, δ varies from -0.440 to 0.440 radians.

The angle, ω, through which Mars rotates on its axis from the start of a Martian day to the moment of sunrise, at point A, is given by $\cos\omega = 0.839\tan\delta$, $0 \le \omega \le \pi$.

(c) (i) Show that the maximum value of $\omega = 1.98$, correct to three significant figures. [3]

(ii) Find the minimum value of ω. [1]

(d) Use your answers to parts (b) and (c) to find

(i) the maximum value of $R(t)$; [2]

(ii) the minimum value of $R(t)$. [1]

(e) Hence show that $a = 1.6$, correct to two significant figures. [2]

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

Let $S(t)$ be the length of time, in hours, from the start of the Martian day until **sunset** at point A on day t. $S(t)$ can be modelled by the function

$$S(t) = 1.5\sin(0.00939t + 2.83) + 18.65\,.$$

The length of time between sunrise and sunset at point A, $L(t)$, can be modelled by the function

$$L(t) = 1.5\sin(0.00939t + 2.83) - 1.6\sin(0.00939t) + d\,.$$

(g) Find the value of d. [2]

Let $f(t) = 1.5\sin(0.00939t + 2.83) - 1.6\sin(0.00939t)$ and hence $L(t) = f(t) + d$.

$f(t)$ can be written in the form $\mathrm{Im}(z_1 - z_2)$, where z_1 and z_2 are complex functions of t.

(h) (i) Write down z_1 and z_2 in exponential form, with a constant modulus. [3]

(ii) Hence or otherwise find an equation for L in the form $L(t) = p\sin(qt + r) + d$, where p, q, r, $d \in \mathbb{R}$. [4]

(iii) Find, in hours, the shortest time from sunrise to sunset at point A that is predicted by this model. [2]

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]

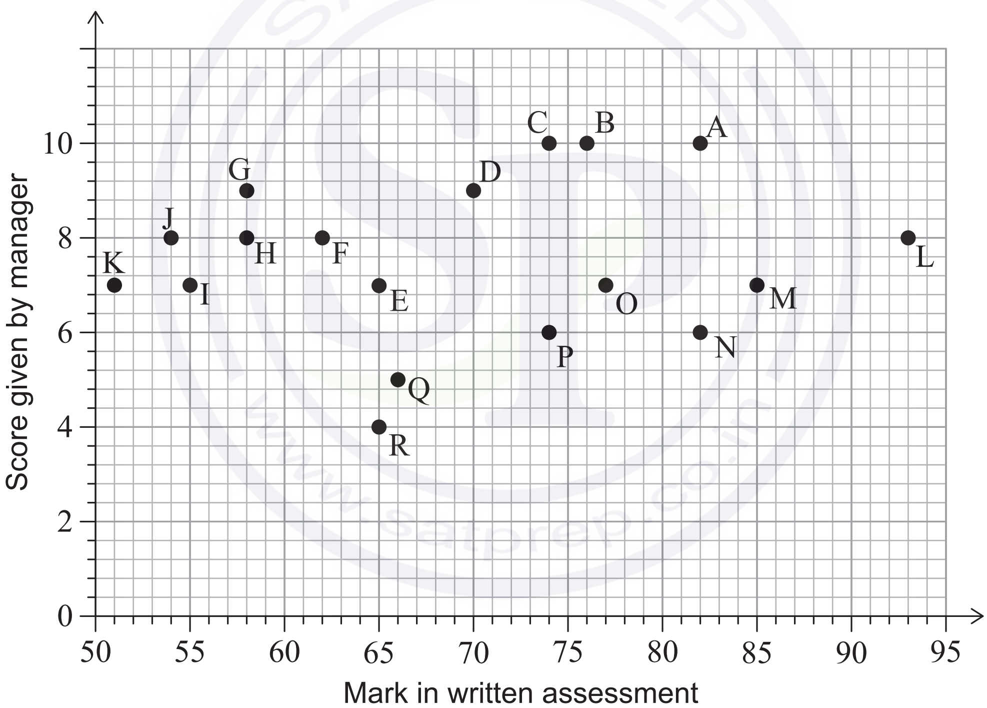
(This question continues on the following page)

(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]

(This question continues on the following page)

(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, $\mathrm{H}_0 : \rho = 0$, against the alternative hypothesis, $\mathrm{H}_1 : \rho > 0$, 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]

References:

Diploma Programme
Programme du diplôme
Programa del Diploma

Mathematics: applications and interpretation
Higher level
Paper 3

Specimen paper

1 hour

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- A graphic display calculator is required for this paper.
- Answer all the questions in the answer booklet provided.
- Unless otherwise stated in the question, all numerical answers should be given exactly or correct to three significant figures.
- A clean copy of the **mathematics: applications and interpretation formula booklet** is required for this paper.
- The maximum mark for this examination paper is **[55 marks]**.

Answer **both** questions in the answer booklet provided. Please start each question on a new page. Full marks are not necessarily awarded for a correct answer with no working. Answers must be supported by working and/or explanations. Solutions found from a graphic display calculator should be supported by suitable working. For example, if graphs are used to find a solution, you should sketch these as part of your answer. Where an answer is incorrect, some marks may be given for a correct method, provided this is shown by written working. You are therefore advised to show all working.

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]

(This question continues on the following page)

(Question 1 continued)

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]

(This question continues on the following page)

(Question 1 continued)

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]

(This question continues on the following page)

(Question 1 continued)

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.

	$(f-p)<-2$	$-2\le(f-p)<0$	$0\le(f-p)<2$	$(f-p)\ge 2$
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]

2. [Maximum mark: 28]

The number of brown squirrels, x, in an area of woodland can be modelled by the following differential equation.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{x}{1000}(2000 - x) \text{, where } x > 0$$

(a) (i) Find the equilibrium population of brown squirrels suggested by this model.

(ii) Explain why the population of squirrels is increasing for values of x less than this value. [3]

One year conservationists notice that some black squirrels are moving into the woodland. The two species of squirrel are in competition for the same food supplies. Let y be the number of black squirrels in the woodland.

Conservationists wish to predict the likely future populations of the two species of squirrels. Research from other areas indicates that when the two populations come into contact the growth can be modelled by the following differential equations, in which t is measured in tens of years.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{x}{1000}(2000 - x - 2y),\ x,\ y \geq 0$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{y}{1000}(3000 - 3x - y),\ x,\ y \geq 0$$

An equilibrium point for the populations occurs when both $\frac{\mathrm{d}x}{\mathrm{d}t} = 0$ and $\frac{\mathrm{d}y}{\mathrm{d}t} = 0$.

(b) (i) Verify that $x = 800$, $y = 600$ is an equilibrium point.

(ii) Find the other three equilibrium points. [6]

(This question continues on the following page)

(Question 2 continued)

When the two populations are small the model can be reduced to the linear system

$$\frac{dx}{dt} = 2x$$
$$\frac{dy}{dt} = 3y\ .$$

(c) (i) By using separation of variables, show that the general solution of $\frac{dx}{dt} = 2x$ is $x = Ae^{2t}$.

(ii) Write down the general solution of $\frac{dy}{dt} = 3y$.

(iii) If both populations contain 10 squirrels at $t = 0$ use the solutions to parts (c) (i) and (ii) to estimate the number of black and brown squirrels when $t = 0.2$. Give your answers to the nearest whole numbers. [7]

For larger populations, the conservationists decide to use Euler's method to find the long-term outcomes for the populations. They will use Euler's method with a step length of 2 years ($t = 0.2$).

(d) (i) Write down the expressions for x_{n+1} and y_{n+1} that the conservationists will use.

(ii) Given that the initial populations are $x = 100$, $y = 100$, find the populations of each species of squirrel when $t = 1$.

(iii) Use further iterations of Euler's method to find the long-term population for each species of squirrel from these initial values.

(iv) Use the same method to find the long-term populations of squirrels when the initial populations are $x = 400$, $y = 100$. [7]

(e) Use Euler's method with step length 0.2 to sketch, on the same axes, the approximate trajectories for the populations with the following initial populations.

(i) $x = 1000$, $y = 1500$

(ii) $x = 1500$, $y = 1000$ [3]

(f) Given that the equilibrium point at $(800, 600)$ is a saddle point, sketch the phase portrait for $x \geq 0$, $y \geq 0$ on the same axes used in part (e). [2]