4.3 Further Correlation & Regression

Question Paper

| Course	DP IB Maths
Section	4. Statistics & Probability
Topic	4.3 Further Correlation & Regression
Difficulty	Hard

Time allowed: 110

Score: /91

Percentage: /100

Question 1a

Jun is investigating the growth patterns of a certain species of slug and measures their thickness, t, and length, l. The results are shown below.

Thickness, t (mm)	9	8	4	3	6	10
Length, l (cm)	5.9	4.1	2.1	1.8	3.0	7.8

a)
Draw a scatter diagram of the data and explain why a line of best fit should not be used for this data.
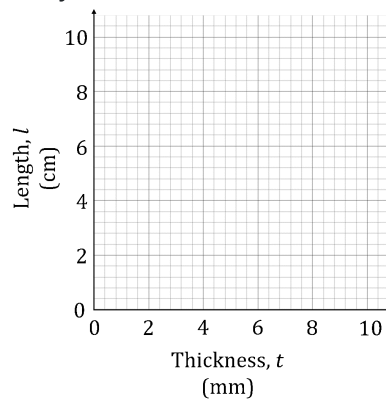
[3 marks]

Question 1b

Jun thinks that an exponential model of the form $l = 2(1.15)^t$ may fit the data, his sister Lily thinks that a quadratic model of the form $l = 0.06t^2 + 1$ is a better fit for the data.

b)
Calculate the sum of the squared residuals, SS_{res}, for

(i)
Jun's model,

(ii)
Lily's model.

[4 marks]

Question 1c

c)

Hence, choose whether Jun's model or Lily's model best fits the data, giving a reason for your answer.

[2 marks]

Question 2a

Wallace wants to model the path of some cylindrical cheeses as they roll down a hill. He measures the vertical distance from the bottom of the hill and the horizontal distance from the point at which the cheese was released at specific points on the path that the cheese takes. These measurements are recorded in the table below.

Horizontal Distance, x (m)	0	1	2	3	4	5	6	7	8
Vertical Distance, y (m)	8.00	7.47	6.08	4.19	2.23	0.23	0.00	0.29	1.46

a)

By first plotting a scatter diagram on your GDC, choose whether a sinusoidal model or a power model would be a better fit for the data.

[2 marks]

Question 2b

b)
Use your graphic display calculator to find the best fit function for the data points.

[2 marks]

Question 2c

c)
Find the coefficient of determination and comment on the closeness of fit to the original data.

[2 marks]

Question 2d

d)
Explain why this model could not be used outside of the range of the data collected.

[1 mark]

Question 3a

A café in the UK begins selling ice creams on the 1st April each year and keeps track of their ice cream sales each month until they stop selling them at the end of October. The results from the year 2021 are shown in the table below.

Month	April	May	June	July	August	Sept	Oct
Sales	82	142	391	516	728	312	64

a)
Suggest a reason why the café owner may choose to use a quadratic function to model the monthly number of ice cream sales.

[2 marks]

Question 3b

b)
Use technology to find the quadratic function of best fit for the data.

[2 marks]

Question 3c

c)
Find the coefficient of determination and interpret the result in context.

[2 marks]

Question 3d

d)
The café owner believes he may have miscalculated the ice cream sales in August. Use the model to find an estimate for the true number of sales of ice creams in August. Comment on the reliability of using the model in this context.

[2 marks]

Question 4a

An ecologist is researching the connection between the mass of different species of rabbits and the spread of their population. She is looking at time taken for a population of 100 to increase to 1000 in four different species of rabbit. The table below shows the average mass of an adult male and the time, in months, for the population to reach 1000.

Rabbit Species	Brush Rabbit	Lionhead	Swamp Rabbit	English Lop
Average mass (kg)	0.7	1.2	2.1	5.2
Time (months)	7	8	11	24

The ecologist believes that the amount of time for a population to reach 1000 can be modelled by an exponential function of the form $T = 5.89\ e^{0.273m}$ where T is the time in months, and m is the mass in kg.

a)
Using the exponential model, find the predicted time taken for each species to reach a population of 1000 and hence the sum of the squared residuals, SS_{res}, for the model.

[3 marks]

Question 4b

b)
Using technology, find the coefficient of determination for
(i)
the exponential model,
(ii)
the least squares regression quadratic model.

[3 marks]

Question 4c

c)

Hence state which function is more appropriate to model the amount of time for a population to reach 1000. Give a reason for your answer.

[2 marks]

Question 4d

d)

State one concern about the reliability of the model.

[1 mark]

Question 5a

As a unicorn moves through the sky its magical sparkling trail becomes more prominent and draws a crowd of elves out to see it. The unicorn flies up high into the sky for a short while, drops down to a height of just above the elves' heads and then comes to its landing place on a cliff top 10 metres above ground level. The horizontal and vertical distances, in metres, of the sparkling trail from the base of the rock that the unicorn started from have been measured for several points and recorded in the table below.

x	10	20	30	40	50	60
y	53	82	116	95	3	10

a)

Find an appropriate cubic model and quartic model for the data, giving all coefficients correct to four significant figures.

[3 marks]

Question 5b

b)

Find the coefficient of determination for both models found in part (a) and comment on the reliability of each model.

[3 marks]

Question 5c

c)

State an alternative type of model that may be more suitable for the data and find

(i)

an equation for this model,

(ii)

the coefficient of determination for this model.

[3 marks]

Question 5d

d)

Comment on why this model is a better fit for the sparkly trail than the initial cubic and quartic models.

[2 marks]

Question 6a

Scientists are collecting information about oxygen levels and temperature in the ocean. They collect data from various different sites. The information is recorded in the table below.

Temperature at surface, T (°C)	13.2	11.1	23.6	15.1	4.2	29.1	19.1
Dissolved Oxygen Content, D (mgL^{-1})	9.1	10.9	8.7	9.0	13.6	8.1	8.8

a)
Use your graphic display calculator to

(i)
find an appropriate quadratic model and logarithmic model for the data,

(ii)
investigate whether a quadratic or a logarithmic model best fits the data, giving reasons for your answer.

[5 marks]

Question 6b

b)
Explain what action the scientists could take to further investigate which model is best.

[1 mark]

Question 6c

c)
For both models found in part (a), find an approximation for the dissolved oxygen content in an ocean of $35°C$ and comment on the validity of your answers.

[3 marks]

Question 7a

A new company believes that their net profit, t months after opening, can be modelled by the equation

$$P(t) = Ate^{bt}, \qquad A, b \in \mathbb{R}$$

Their net profit at different times over the first year is given in the table below.

t (months)	2	5	10	12
P ($1000)	0.2	2.5	51.6	201.9

a)
Given $A = 0.04$ and $b = 0.5$, calculate the sum of the squared residuals, SS_{res} for this model and comment on its suitability.

[4 marks]

Question 7b

b)
Show that a second model for the net profit over time can be given as
$$\ln P = 0.6764t - 2.764$$

[4 marks]

Question 7c

c)
Choosing the model from part (a) or (b) that you believe to fit the data best, find an estimate for the net profit gained by the company after 20 months.

[2 marks]

Question 7d

d)
Comment on the reliability of the answer found in part (c), giving a reason for your answer.

[1 mark]

Question 8a

A biologist is researching a connection between the mass of an animal, M kg, and its expected lifespan, L years. The biologist suggests that there exists a relationship of the form $L = AM^B$, where A and B are constants to be found.

a)
Show that the relationship can be rewritten using logarithms as
$$\log L = \log A + B \log M$$

[3 marks]

Question 8b

Using data from a wide range of animals, when $y = \log L$ is plotted against $x = \log M$ on a scatter diagram there seems to be a strong positive correlation. When the regression line of y on x is calculated, the equation is found to be $y = 0.18x + 0.98$.

b)
By relating the equation of the regression line to the equation found in (a), or otherwise, find the constants A and B correct to 2 decimal places where appropriate

[2 marks]

Question 8c

c)
Hence, predict the lifespan of a horse with a mass of 600 kg to the nearest year.

[1 mark]

Question 8d

The biologist concludes the research by suggesting that one way to increase your lifespan is to increase your mass.

d)
Explain, based on these data, why the biologist may be incorrect.

[1 mark]

Question 9a

A virologist is studying the growth rate of a particular type of virus when attached to a particular host cell. They record the number of cells of the virus and the time in minutes that has elapsed since the virus attached to the host cell. The results are recorded in the table below.

Time (t hours)	0	1	2	3	4	5
Number of cells (n)	1	22	604	11270	125242	1007518

The virologist wants to linearise the data. They take logarithms of the number of cells for $1 \leq t \leq 5$ and draw a semi-log graph of their calculated data.

a)
Draw a semi-log graph of the calculated data $y = \ln n$, against time, t, for $1 \leq t \leq 5$.

[4 marks]

Question 9b

b)
Find the equation of the regression line of y on t, giving your answer in the form $y = at + b$, where a and b are constants to be found.

[2 marks]

Question 9c

c)
By substituting for into the answer for part (b), find a model for t and n in the form $n = Ae^{ct}$ where A and c are constants to be found.

[3 marks]

Question 9d

d)
Find an estimate for the number of virus cells present after 3.5 hours, comment on the reliability of this estimate.

[2 marks]

Question 10a

Olivia is modelling a new bowl for her pottery shop. She models the outline of one side of the bowl on a 2D Cartesian coordinate grid and plans to rotate the design 360° about the y-axis.
The coordinates Olivia uses to plot the cross-section are given in the table below.

Point	A	B	C	D	E	F	G	H	I
x	5	6	8	10	12	14	16	18	18
y	0	2	4	5	6	6	8	10	12

Point A is connected to the origin and point I is connected to the point $(0, 12)$ with a straight, horizontal line.

a)
Olivia initially models all of the points using a cubic curve.
(i)
Find the equation of the least squares regression cubic curve for all nine points.
(ii)
Find the coefficient of determination for the cubic model.
(iii)
Explain why this model is not a good model for all of the points and state which points Olivia should use a different model for.
Use mathematical reasoning to validate your argument.

[5 marks]

Question 10b

b)
Olivia decides instead to use two different linear models between points A and C and C and E and then use a quadratic model to connect points E, F, G, H and I. Find the equation of this quadratic model and write down a problem with using this model instead.

[3 marks]

Question 10c

c)

State which model Olivia should use and any limitations to this model.

[1 mark]