

IB Maths: AI HL

Further Correlation & Regression

Topic Questions

These practice questions can be used by students and teachers and is Suitable for IB Maths AI HL Topic Questions

Course	IB Maths
Section	4. Statistics & Probability
Topic	4.3 Further Correlation & Regression
Difficulty	Medium

Level: IB Maths

Subject: IB Maths AI HL

Board: IB Maths

Topic: Further Correlation & Regression

Question 1

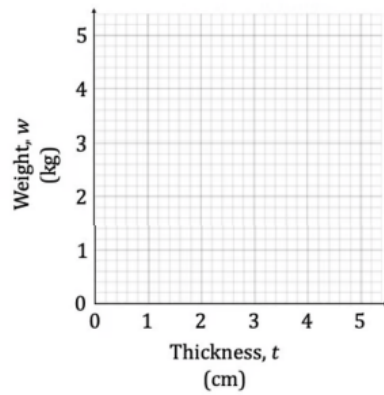
Taro discovers a group of rocks and decides to measure their thickness, t , and weight, w . The results are shown below.

Thickness, t (cm)	2.2	4.5	3.2	2.8	3.7
Weight, w (kg)	2.6	3.6	2.7	3.0	2.9

Taro draws a line of best fit through the data points using the equation $w = \frac{2}{5}t + \frac{8}{5}$.

(a)

Draw a scatter diagram of the data and sketch the given line of best fit on the same axes.



[3 marks]

(b)

Calculate the residual of each point.

[3 marks]

(c)

Hence find the sum of the squared residuals, SS_{res} .

[1 mark]

(d)

Comment on the fit of the model to the data.

[2 marks]

Question 2

Syafiqah wants to model the path of the water from a water fountain. She measures the horizontal and vertical distances of specific points on the path that the water takes with respect to the base of the water spout. These measurements are recorded in the table below.

Horizontal Distance, x (m)	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Vertical Distance, y (m)	1.32	1.47	1.50	1.56	1.39	1.33	0.98	0.63	0.00

(a)

State an appropriate type of function to model the path way of the water.

[1 mark]

(b)

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

[2 marks]

(c)

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

[2 marks]

(d)

State the maximum height that the water reaches according to the model and the horizontal distance of this point from the base of the water spout.

[3 marks]

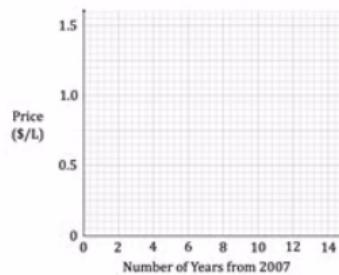
Question 3

The average price for cooking oil in USD per litre was recorded each year from 2007 until 2021 and the results put in the table below. Number of years from 2007.

Year	2007	2009	2011	2013	2015	2017	2019	2021
Price (USD/L)	0.78	1.32	1.31	1.22	0.99	0.85	0.84	1.24

(a)

Draw a scatter graph of the data points on the axes provided below, with the price in USD per litre against the number of years from 2007.



[3 marks]

(b)

State the type of function that would be appropriate to model the data.

[1 mark]

(c)

Using your graphic display calculator find the function of best fit for the data.

[2 marks]

(d)

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

[2 marks]

(e)

Sketch the model function on top of the scatter graph in part (a) and comment on the closeness of fit to the original data.

[3 marks]

Question 4

A specialist in infectious diseases is investigating the spread of a new disease. She is looking at the number of infections in a community based on the time, in days, since the first recorded case in the area.

She has focused on three sites and recorded the following information.

Site	1	2	3
Time since first recorded case, t	35	180	166
Number of infections, N	9	411	247

The scientist believes that the number of infections in a population can be modelled by an exponential function of the form

$$N = A(1.2)^{bt}, \text{ where } A \text{ and } b \text{ are constants.}$$

(a)

Given that $A = 0.6$ and $b = 0.2$, find the predicted number of infections and hence the residuals for each site.

[3 marks]

A second model proposes that $A = 2.5$ and $b = 0.15$.

(b)

Find the residuals for each site for the second model.

[2 marks]

The scientist will use the model that has the lowest value for the sum of the squares of the residuals.

(c)

Determine which model the scientist should use.

[3 marks]

(d)

State one concern about the reliability of the model.

[1 mark]

Question 5

A unicorn leaves a magical sparkling trail behind him when he jumps off a rock that is 3 metres above ground level and flies away into the sky. The horizontal and vertical distances in metres from the base of the rock that the unicorn started from have been measured for several points in the sparkling trail and recorded in the table below.

x	0	2	4	6	7	10
y	3	3	4.2	9.5	15	53.3

(a)

Find an appropriate quadratic model for the data.

[2 marks]

(b)

Find the coefficient of determination for the quadratic model and comment on the closeness of fit to the data.

[3 marks]

(c)

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

[1 mark]

(d)
Find an equation for the alternative model.

[2 marks]

(e)
Comment on why this model is a better fit for the sparkly trail than the initial quadratic.

[2 marks]

Question 6

Grainne records the displacement, in centimetres, of a particle from a fixed point O over the course of 9 seconds. She wishes to find an equation that will model the movement of the particle.

The coordinates of some of the data points of the particle are shown in the table below.

Time, t	0.9	1.2	2.7	3.9	4.2	5.1	5.7	7.7	8.3
Displacement, s	-5.9	-7.0	-8.2	-4.5	-3.0	-3.2	-5.8	-7.9	-6.0

(a)
State why a quadratic would not be a good fit for this model. Give a reason for your answer.

[2 marks]

(b)
State a type of function that would be appropriate to model the particle's movement.

[1 mark]

(c)
Find an equation of the least squares regression for the type of function stated in part (b).

[2 marks]

(d)
Given that the displacement of the particle from point O is -4.7 cm, find the time(s) at which this occurs, for $0 \leq t \leq 9$.

[3 marks]

Question 7

A professional skydiver is attempting to perform the highest skydive without the use of specialist breathing equipment. A sensor has been released from a height of 12,000 m to record the atmospheric pressure at different heights above sea level. The data from the sensor is recorded in the table below.

Height above sea level, h (m)	0	2885	5476	7732	9640	12000
Atmospheric pressure, p (Pa)	1013	766	483	350	301	215

An exponential model of the form $p = a(b)^h$ is proposed to model the data, where a , b are constants to be found.

(a)

Find the values of a and b .

[3 marks]

(b)

Comment on any limits there might be for the domain of this model.

[2 marks]

It is agreed that an oxygen supply should be used when the atmospheric pressure is less than 365 Pa.

(c)

State whether a jump from a height of 8000 m could be attempted safely without an oxygen supply.

[3 marks]

(d)

Explain what action could be taken to increase the reliability of the model.

[1 mark]

Question 8

A digital artist has taken a photograph of a building and wishes to render it digitally. The photograph is imported to graphing software and the following points that follow the line of the roof are plotted.

x	0	0.7	1.8	2.3	3.3	4.0	4.5
y	1.2	1.5	1.7	2.9	6.5	8.0	13.8

One possible model for the roof is a cubic.

Another possible model for the roof is an exponential function, $y = ab^x$, where $a, b \in \mathbb{R}$.

(a)

Find the equation of the least squares regression curve for

(i)

the cubic model

(ii)

the exponential model.

[4 marks]

(b)

Find the coefficient of determination for

(i)

the cubic model

(ii)

the exponential model.

[3 marks]

(c)

Hence state which function is more appropriate to model the roof line of the building. Give a reason for your answer.

[2 marks]

(d)

Using the model that was chosen in part (b), find the height of the roof when $x = 9.3$.

[2 marks]

Question 9

Mo is looking at the growth rate of a particular type of bacteria. She records the number of cells of the bacteria and the time in minutes that has elapsed since the start of her observation. The results are recorded in the table below.

Time (minutes)	0	2	4	6	8	10	12	14
Number of cells	1	3	16	112	2667	6310	210863	914113

(a)

Express the number of cells as a logarithm. Give each value to 3 dp.

[3 marks]

Mo is not sure if a linear model or a quadratic model will fit the data best.

(b)

Find the coefficient of determination for

(i)

a linear model

(ii)

a quadratic model

[3 marks]

(c)

Comment on which model from part (c) best fits the graph.

[1 mark]

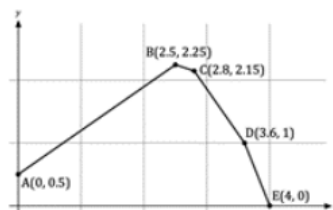
(d)

Find the equation of the least squares regression curve for the model specified in part (c).

[2 marks]

Question 10

Amelie wants to create a scale model of the cross section of a hill. She has measured the horizontal and vertical distances of several points on the hill from a fixed point O. These points can be seen in the diagram below.



Amelie decides to model the first section as a straight line and the second section as a quadratic with equation $y = -0.9x^2 + 4.5x - 3.3$.

Find the equation of the straight line passing through points A and B.

[2 marks]

(b)

Find the residuals for each point B, C, D and E for the quadratic model.

[4 marks]

(d)

By examining the sum of the squares of the residuals, show that the model found in part (c) is a better fit for the data.

[4 marks]

Question 11

The number of active profiles, N , a new social media platform has t months after launching are shown in the table below.

t	3	5	8	12	15	24
N	542	1122	2345	4254	8621	22 324
$\log N$						

(a) Complete the third row of the table above.

[2 marks]

b) Draw a scatter diagram of $\log N$ versus t .

[3 marks]

The regression line of $\log N$ on t can be written in the form $\log N = a + bt$.

(c) Find the values of a and b .

[3 marks]

d)

Calculate the Pearson's product-moment correlation coefficient, r , between $\log N$ and t and comment on the validity of the regression line found in part (c).

[3 marks]

e) Based on part (d) suggest a suitable type of regression model for N on t .

[1 mark]