IB Math AI HL — Topic 4

Non-linear Regression & Logarithmic Scales

Least Squares Regression Curves · Coefficient of Determination · Logarithmic Scales · Linearising using Logarithms

Difficulty	EASY	MEDIUM	HARD
Questions	10	10	10
Total Marks	30	40	50

Scan me for help

Name:____________________________ Date:______________

Score:_________/120

Instructions: Answer all questions. Show all working. Give answers correct to 3 significant figures unless stated otherwise. A GDC is required.

Key Formulae

- **Coefficient of determination:** R^2 (from GDC), where $0 \leq R^2 \leq 1$. A value close to 1 indicates a strong fit. For linear regression, $R^2 = r^2$ where r is the PMCC.
- **Power model** $y = ax^b$: take $\log_{10}$ of both sides $\Rightarrow$ $\log y = \log a + b \log x$. Plot $\log y$ vs $\log x$ (log-log); gradient $= b$, intercept $= \log a$.
- **Exponential model** $y = ab^x$: take $\log_{10}$ of both sides $\Rightarrow$ $\log y = \log a + x \log b$. Plot $\log y$ vs x (semi-log); gradient $= \log b$, intercept $= \log a$.
- **Back-transforming:** $a = 10^{\text{intercept}}$; $b = 10^{\text{gradient}}$ (for exponential) or $b =$ gradient (for power).
- **Richter scale:** $M = \log_{10}\left(\frac{I}{I_0}\right)$; **Decibel:** $L = 10\log_{10}\left(\frac{I}{I_0}\right)$; **pH:** $\text{pH} = -\log_{10}[\text{H}^+]$.
- **Ratio from scale difference:** If $M_2 - M_1 = d$, then $\frac{I_2}{I_1} = 10^d$ (Richter) or $10^{d/10}$ (decibel).
- **SSres** (residual sum of squares): smaller SSres $\Rightarrow$ better fit; compare models fitted to the same data.

GDC Tips

- **Non-linear regression on GDC:** `STAT` $\rightarrow$ `CALC` $\rightarrow$ choose model (ExpReg, PwrReg, QuadReg, LnReg, SinReg). Always check R^2 reported alongside.
- **Storing lists:** Enter raw data in `L1`/`L2`; for linearisation enter $\log x$ in `L3` and $\log y$ in `L4`, then run LinReg on L3, L4.
- **Comparing models:** Run each regression, record R^2 and SSres. A higher R^2 (or lower SSres) on the *original* scale indicates a better model.

- **Reading R^2 from GDC:** On TI-84, enable DiagnosticOn (`2nd` $\rightarrow$ `CATALOG`). On Casio fx-CG50: `STAT` $\rightarrow$ `CALC` $\rightarrow$ select regression type.

Section A — Easy EASY (10 questions × 3 marks = 30 marks)

Scan me for help

1. EASY Calculator **[3 marks]**

The table below gives the number of active users N (thousands) of a new social media platform, recorded at the end of each month t.

t (months)	1	2	3	4	5
N (thousands)	4.2	9.8	22.6	52.4	121.5

An exponential regression model of the form $N = ab^t$ is fitted to the data.

(a) Write down the value of a and the value of b. [2]

(b) Write down the coefficient of determination, R^2. [1]

2. EASY Calculator **[3 marks]**

The sound intensity level L in decibels (dB) is given by

$$L = 10\log_{10}\left(\frac{I}{I_0}\right),$$

where I is the intensity of the sound in W m^{-2} and $I_0 = 1 \times 10^{-12}$ W m^{-2}.
A pneumatic drill produces a sound intensity of $I = 3.16 \times 10^{-3}$ W m^{-2}.

(a) Find the sound intensity level L of the drill. [2]

(b) State one limitation of using a logarithmic scale to report sound intensity. [1]

3. EASY Calculator **[3 marks]**

A researcher suspects that the variables x and y satisfy a relationship of the form $y = ab^x$, where $a > 0$ and $b > 0$.

(a) Show that taking $\log_{10}$ of both sides gives a linear relationship between $\log_{10} y$ and x. [1]

(b) State the gradient and the y-intercept of this linear relationship in terms of a and/or b. [2]

4. EASY Calculator **[3 marks]**

Amara records the area A (cm^2) of a fungal colony on each day d after it was first observed. She fits two regression models to her data using a GDC and obtains:

Model	Equation	R^2
Quadratic	$A = 0.31d^2 + 0.45d + 1.2$	0.962
Exponential	$A = 1.85 \times 1.74^d$	0.997

(a) State which model provides a better fit to the data. Justify your answer. [2]

(b) Use the better model to predict the area of the colony on day 7. [1]

5. EASY Calculator **[3 marks]**

The pH of a solution is defined as pH $= -\log_{10}[\text{H}^+]$, where $[\text{H}^+]$ is the hydrogen ion concentration in mol L^{-1}.
A citric acid solution has pH $= 2.3$.

(a) Find the hydrogen ion concentration $[\text{H}^+]$. Give your answer in the form $a \times 10^k$ where $1 \leq a < 10$. [2]

(b) A second solution has pH $= 5.3$. Find the ratio $\dfrac{[\text{H}^+]_{\text{acid}}}{[\text{H}^+]_{\text{second}}}$. [1]

6. EASY Calculator **[3 marks]**

Data for variables x and y are plotted on a semi-logarithmic graph (axes: x horizontal, $\log_{10} y$ vertical). The plotted points lie on the straight line

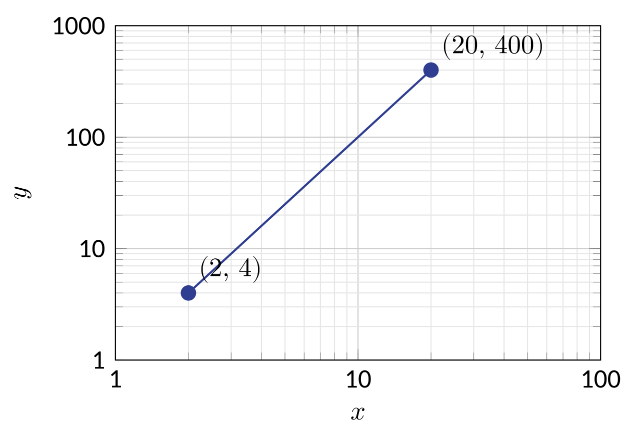
$$\log_{10} y = 0.15x + 1.8.$$

(a) Write down the value of $\log_{10} y$ when $x = 0$. [1]

(b) Hence find the value of a and the value of b in the model $y = ab^x$. [2]

7. EASY Calculator **[3 marks]**

The diagram below shows two data points plotted on a log-log graph (both axes are $\log_{10}$-scaled). The points lie on a straight line.



(a) Find the gradient of the line on the log-log graph. [1]

(b) Hence find the equation connecting x and y in the form $y = ax^b$. Your answer should not include logarithms. [2]

8. EASY Calculator **[3 marks]**

A power regression model $y = 3.2x^{1.7}$ is fitted to a dataset. The coefficient of determination for this model is $R^2 = 0.924$.

(a) Interpret the value of $R^2 = 0.924$ in context. [1]

(b) A linear regression model fitted to the same data gives $R^2 = 0.861$. State, with a reason, which model fits the data better. [2]

9. EASY Calculator **[3 marks]**

A radioactive substance decays according to the model $M = 240 \times (0.87)^t$, where M is the mass in grams and t is the time in hours.

(a) Write down the initial mass of the substance. [1]

(b) Find the time at which the mass is half the initial value. Give your answer correct to 3 significant figures. [2]

10. EASY Calculator **[3 marks]**

The Richter scale magnitude M of an earthquake is defined as

$$M = \log_{10}\left(\frac{I}{I_0}\right),$$

where I is the intensity of the earthquake and I_0 is a reference intensity.
An earthquake in region A has magnitude 4.5. An earthquake in region B has magnitude 6.5.

(a) Find the intensity ratio $\frac{I_B}{I_A}$. [2]

(b) State what this ratio tells you about the two earthquakes. [1]

Section B — Medium **MEDIUM** (10 questions × 4 marks = 40 marks)

Scan me for help

11. **MEDIUM** **Calculator** **[4 marks]**

The table gives the concentration C (mg L^{-1}) of a drug in a patient's bloodstream at time t (hours) after administration.

t (hours)	1	2	3	4	5
C (mg L^{-1})	18.4	13.2	9.5	6.8	4.9

(a) Using a GDC, find the exponential regression model $C = ab^t$. [2]

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

(c) Use the model to predict the concentration at $t = 8$ hours. Comment on the reliability of this prediction. [1]

12. **MEDIUM** **Calculator** **[4 marks]**

A marine biologist records the mass M (g) of a growing coral fragment over several weeks w. She suspects the relationship follows $M = ab^w$.

She transforms the data by computing $Y = \log_{10} M$ and performs linear regression of Y on w, obtaining:

$$Y = 0.082w + 0.699.$$

(a) Find the values of a and b in the model $M = ab^w$. [2]

(b) Hence predict the mass of the coral at week 12. [1]

(c) State one assumption made in applying this model beyond week 12. [1]

13. MEDIUM Calculator **[4 marks]**

The table gives the period T (seconds) of a pendulum and the length L (cm) of the pendulum string.

L (cm)	25	50	75	100	150
T (s)	1.00	1.42	1.74	2.01	2.46

(a) Explain why a log-log linearisation is appropriate for data of the form $T = aL^b$. [1]

(b) Using a GDC, perform a power regression to find the values of a and b. Give each value correct to 3 significant figures. [2]

(c) Write down the value of R^2 and comment on whether the power model is appropriate. [1]

14. MEDIUM Calculator **[4 marks]**

The sound intensity level is $L = 10\log_{10}\left(\dfrac{I}{I_0}\right)$ dB, where $I_0 = 1 \times 10^{-12}$ W m^{-2}.

A concert hall measures a noise level of 94 dB at the mixing desk. After fitting acoustic panels, the noise level falls to 79 dB.

(a) Find the sound intensity I at the mixing desk before the panels were fitted. Give your answer in the form $a \times 10^k$ W m^{-2}. [2]

(b) Find the ratio $\dfrac{I_{\text{before}}}{I_{\text{after}}}$. [2]

15. MEDIUM Calculator **[4 marks]**

The table gives the daily revenue R (USD) of a market stall during its first six days of trading.

d (day)	1	2	3	4	5	6
R (USD)	85	140	198	249	285	310

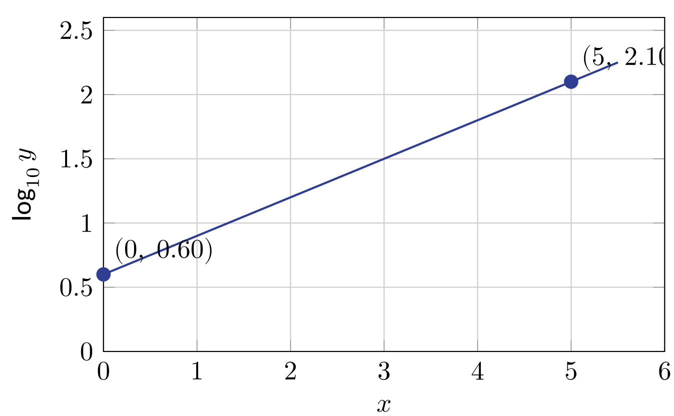
(a) Using a GDC, find the equation of the quadratic regression model $R = ad^2 + bd + c$. Write the coefficients correct to 3 significant figures. [2]

(b) The coefficient of determination for the quadratic model is $R^2 = 0.998$. A logarithmic regression model of the form $R = p \ln d + q$ gives $R^2 = 0.969$. State, with a reason, which model is more appropriate. [1]

(c) Use the quadratic model to estimate the revenue on day 10 and comment on the reliability of this estimate. [1]

16. MEDIUM Calculator **[4 marks]**

The diagram shows a semi-logarithmic graph of data points for x and $\log_{10} y$. The points lie exactly on the line shown, which passes through $(0, 0.60)$ and $(5, 2.10)$.



(a) Write down the equation of the line in the form $\log_{10} y = mx + c$. [1]

(b) Find the values of a and b such that $y = a \cdot b^x$. [2]

(c) Hence find the value of y when $x = 3$. [1]

17. **MEDIUM** **Calculator** **[4 marks]**

A sports scientist records the maximum speed v (m s^{-1}) achieved by an athlete as a function of training hours h per week. The data are shown below.

h (hours per week)	4	8	12	16
v (m s^{-1})	6.1	7.3	8.0	8.6

The scientist believes the relationship follows a power law $v = ah^b$.

(a) Complete the table of transformed data:

$\log_{10} h$				
$\log_{10} v$				

[1]

(b) Perform a linear regression on the transformed data to find a and b. Give each value correct to 3 significant figures. [3]

18. MEDIUM Calculator **[4 marks]**

A function is defined as $f(x) = \log_{10}(x+3) + \log_{10}(x-1)$ for $x > 1$.

(a) Write $f(x)$ as a single logarithm in the form $\log_{10} g(x)$. [1]

(b) Solve $f(x) = 1$, giving all solutions in exact form. Discard any values outside the domain. [3]

19. MEDIUM Calculator **[4 marks]**

The table gives the metabolic rate B (watts) and the body mass m (kg) of six mammal species.

m (kg)	0.02	0.30	3.5	25	70	500
B (W)	0.04	0.31	2.0	9.5	21	94

(a) Use a GDC to fit a power regression model $B = am^b$. [2]

(b) Write down R^2 and interpret it in context. [1]

(c) Interpret the value of the exponent b in the context of this biological data. [1]

20. MEDIUM Calculator **[4 marks]**

The table gives the number of bacteria N (millions) in a culture at time t (hours).

t (hours)	0	1	2	3	4
N (millions)	2.0	3.5	6.3	11.1	19.8

An exponential model $N = ab^t$ is proposed.

(a) Using the GDC, find a and b and write down R^2. [2]

(b) Calculate the residual for the data point $(t = 2,\ N = 6.3)$. State the meaning of this residual. [2]

Section C — Hard **HARD** (10 questions × 5 marks = 50 marks)

Scan me for help

21. **HARD** **Calculator** **[5 marks]**

The table gives the population P (thousands) of a coastal town t years after 2010.

t (years)	2	5	8	11	14	17
P (thousands)	38.2	44.6	52.1	60.9	71.2	83.1

(a) Using a GDC, fit an exponential model $P = ab^t$ and write down R^2. [2]

(b) By taking $\log_{10}$ of both sides of $P = ab^t$, linearise the model. Using linear regression on $\log_{10} P$ and t, obtain the gradient and intercept, and hence find a and b to 3 significant figures. [2]

(c) Determine whether the town's population will exceed 120 thousand within 25 years of 2010. Justify your answer. [1]

22. HARD Calculator **[5 marks]**

Data for two variables x and y are plotted on a log-log graph. The plotted points lie exactly on a straight line.

Given that the line passes through $(3,\ 45)$ and $(12,\ 720)$ on the log-log scale (i.e. these are the actual $(x,\ y)$ values, not the log values):

(a) Show that the gradient of the line on the log-log graph is 2. [2]

(b) Hence find the exact values of a and b in the model $y = ax^b$. [2]

(c) A third data point is $(k,\ 5000)$. Find the value of k, giving your answer correct to 3 significant figures. [1]

23. HARD Calculator **[5 marks]**

Two sound sources, P and Q, are measured in a recording studio. Let I_P and I_Q be their intensities in W m^{-2}.

It is known that:

- the decibel level of P is 72 dB;
- the decibel level of Q exceeds that of P by 15 dB;
- $I_0 = 1 \times 10^{-12}$ W m^{-2}.

(a) Find I_P. [2]

(b) Find I_Q. [1]

(c) Both sources play simultaneously. Assuming the intensities add, find the combined decibel level. Give your answer correct to 1 decimal place. [2]

24. **HARD** **Calculator** **[5 marks]**

A data scientist models the monthly downloads D (thousands) of an app using $D = 5 \cdot b^t$, where t is the number of months since launch.

The following partial data are available:

t (months)	1	2	3	4	5
D (thousands)	9.3	18.1	34.7	67.2	128.4

(a) The scientist linearises by taking $\log_{10} D$ and regressing on t. Show that the gradient of the linearised regression line equals $\log_{10} b$. [2]

(b) Using a GDC, perform the linearised regression and hence find b. Give your answer correct to 3 significant figures. [2]

(c) The app is considered a commercial success if D first exceeds 500 thousand in at most 8 months. Determine whether this criterion is met. [1]

25. **HARD** **Calculator** **[5 marks]**

The table gives the resistance R (ohms) of a wire as a function of temperature θ (°C).

θ (°C)	20	40	60	80	100	120
R (ohms)	12.8	14.6	16.5	18.7	21.0	23.7

(a) Using a GDC, fit an exponential model $R = a \cdot b^{\theta}$ and record R^2. [1]

(b) Linearise by computing $Y = \log_{10} R$ and perform linear regression of Y on θ. Find a and b from the linearised regression line. [3]

(c) The two methods in (a) and (b) may give slightly different values of a and b. Explain why this is so. [1]

26. **HARD** **Calculator** **[5 marks]**

Two data points lie on a log-log straight line: $(x_1, y_1) = (4, 8)$ and $(x_2, y_2) = (16, k)$. The relationship has the form $y = ax^b$.

(a) Show that $b = \dfrac{\log_{10} k - \log_{10} 8}{\log_{10} 4}$. [2]

(b) Given that $b = 1.5$, find the exact value of k and the exact value of a. [2]

(c) A third point $(64, m)$ lies on the same line. Find m. [1]

27. **HARD** **Calculator** **[5 marks]**

The table gives the braking distance d (m) of a car travelling at speed v (km h^{-1}).

v (km h^{-1})	30	50	70	90	110
d (m)	8.2	22.8	44.6	73.8	110.0

(a) Using a GDC, fit each of the following models to the data:

- Power model: $d = av^b$
- Quadratic model: $d = pv^2 + qv + r$

Write down R^2 for each model. [2]

(b) Physics predicts that braking distance is proportional to v^2 (i.e. $b = 2$). State, using both R^2 and this physical argument, which model you would choose. Justify your answer fully. [2]

(c) Using the power model, estimate the braking distance at $v = 130$ km h^{-1}. State one reason why this estimate may be unreliable. [1]

28. HARD Calculator **[5 marks]**

Seismologists record the aftershock intensities of an earthquake. They model the magnitude M of the nth aftershock by

$$M(n) = 5.8 \times (0.78)^n, \quad n = 1, 2, 3, \ldots$$

where M is measured on the Richter scale.

(a) Find the magnitude of the third aftershock, correct to 3 significant figures. [1]

(b) Find the value of n for which the magnitude first falls below 2.0. Show your algebraic working. [2]

(c) The intensity ratio between the first and fifth aftershocks is $\dfrac{I_1}{I_5}$. Show that $\dfrac{I_1}{I_5} = 10^{M_1 - M_5}$ and find this ratio correct to 3 significant figures. [2]

29. HARD Calculator **[5 marks]**

A market analyst models the price P (USD) of a rare mineral as a function of scarcity index s using a power model $P = ks^c$, where k and c are constants.

Two data points are known exactly:

$$(s_1,\ P_1) = (10,\ 3200) \quad \text{and} \quad (s_2,\ P_2) = (40,\ 25600).$$

(a) Taking $\log_{10}$ of both sides, show that $c = \dfrac{\log_{10} 8}{\log_{10} 4}$. [2]

(b) Hence find the exact value of c. [1]

(c) Find the exact value of k. [1]

(d) A new deposit is found with scarcity index $s = 25$. Find the predicted price per unit. [1]

30. **HARD** **Calculator** **[5 marks]**

An ecologist observes that the number of bird species S on islands of area A (km^2) follows a power law $S = cA^z$. The following data are recorded.

A (km^2)	1	4	16	64	256	1024
S (species)	12	20	33	55	91	152

(a) Using log-log linearisation and a GDC, find c and z. Give each value correct to 3 significant figures. [2]

(b) Write down R^2 and comment on the suitability of the model. [1]

(c) The ecologist claims that a new island of area 100 km^2 will support at least 40 species. Use your model to determine whether this claim is supported. [1]

(d) State one limitation of extrapolating this model to very large islands. [1]

Worked Solutions — Official Mark-Scheme Style

M = Method mark A = Accuracy mark R = Reasoning mark AG = Answer Given (no mark) (M1) / (A1) = implied mark (GDC step)

Q1 — Exponential Regression: Users [EASY]

Exponential regression $N = ab^t$ on GDC using data: **(M1)**

(a) $a = 1.83\ldots \approx 1.83; \quad b = 2.31\ldots \approx 2.31$ **A1A1**

Note: Accept $a = 1.82$–1.84, $b = 2.30$–2.32. *Award A1 for each value.*

(b) $R^2 = 0.9998\ldots \approx 0.9998$ **A1**

Note: Accept values in range $[0.999, 1.000]$.

Q2 — Decibel Scale [EASY]

(a) $L = 10\log_{10}\left(\dfrac{3.16 \times 10^{-3}}{1 \times 10^{-12}}\right)$ **M1**

$= 10\log_{10}(3.16 \times 10^{9}) = 10 \times 9.4997\ldots = 95.0$ dB **A1**

(b) Any one reasonable limitation, e.g. "it is not possible to distinguish between intensities of similar magnitude", or "a doubling of intensity produces only a 3 dB increase, which may not convey the actual change to non-specialists." **R1**

Q3 — Identify Linearised Form [EASY]

(a) Taking $\log_{10}$ of $y = ab^x$: **M1**

$\log_{10} y = \log_{10} a + x\log_{10} b$ (AG) **AG**

This is of the form $Y = Mx + C$ with $Y = \log_{10} y$, so it is linear.

(b) Gradient $= \log_{10} b$; $\quad y$-intercept $= \log_{10} a$. **A1A1**

Q4 — Model Comparison by R2 [EASY]

(a) The exponential model is a better fit. **A1**

Reason: Its $R^2 = 0.997$ is closer to 1 than the quadratic's $R^2 = 0.962$, indicating a higher proportion of variance explained. **R1**

(b) Using $A = 1.85 \times 1.74^7$: **(M1)**

$A = 1.85 \times 1.74^7 = 1.85 \times 35.80\ldots = 66.2\ldots \approx 66.2$ cm^2 **A1**

Note: FT from their better model in (a).

Q5 — pH Inversion [EASY]

(a) $2.3 = -\log_{10}[\text{H}^+]$ **M1**

$[\text{H}^+] = 10^{-2.3} = 5.01 \times 10^{-3}$ mol L^{-1} **A1**

(b) $[\text{H}^+]_{\text{second}} = 10^{-5.3}$

Ratio $= \dfrac{10^{-2.3}}{10^{-5.3}} = 10^3 = 1000$ **A1**

Q6 — Back-Transform from Semi-Log Line [EASY]

(a) When $x = 0$: $\log_{10} y = 0.15(0) + 1.8 = 1.8$ **A1**

(b) From intercept: $\log_{10} a = 1.8 \Rightarrow a = 10^{1.8} = 63.1\ldots \approx 63.1$ **A1**

From gradient: $\log_{10} b = 0.15 \Rightarrow b = 10^{0.15} = 1.41\ldots \approx 1.41$ **A1**

Note: Award A1 for each correctly back-transformed value.

Q7 — Log-Log Graph: Power Model [EASY]

(a) On the log-log axes, gradient $= \dfrac{\log_{10} 400 - \log_{10} 4}{\log_{10} 20 - \log_{10} 2}$ **M1**

$= \dfrac{2.602\ldots - 0.602\ldots}{1.301\ldots - 0.301\ldots} = \dfrac{2.000}{1.000} = 2$ **A1**

(b) So $b = 2$. Using point $(2, 4)$: $4 = a \cdot 2^2 \Rightarrow a = 1$. **M1**

$y = x^2$ **A1**

Note: Accept $y = 1 \cdot x^2$.

Q8 — Interpret R2 and Model Choice [EASY]

(a) $R^2 = 0.924$ means that 92.4% of the variation in y is explained by the power model $y = 3.2x^{1.7}$. **R1**

(b) The power model is a better fit. **A1**

Reason: $R^2 = 0.924 > 0.861$, so the power model explains a greater proportion of the variance in the data than the linear model. **R1**

Q9 — Half-Life from Exponential Decay [EASY]

(a) Initial mass $= 240$ g **A1**

(b) Solve $240 \times (0.87)^t = 120$: **M1**

$(0.87)^t = 0.5 \Rightarrow t = \dfrac{\ln 0.5}{\ln 0.87} = \dfrac{-0.6931\ldots}{-0.13926\ldots} = 4.977\ldots \approx 4.98$ hours **A1**

Q10 — Richter Scale Ratio [EASY]

(a) $M_B - M_A = 6.5 - 4.5 = 2.0$ **(M1)**

$\dfrac{I_B}{I_A} = 10^{2.0} = 100$ **A1**

(b) Earthquake B has an intensity 100 times greater than earthquake A. (It releases 100 times more seismic energy per unit area.) **R1**

Q11 — Exponential Regression: Drug Concentration [MEDIUM]

(a) Exponential regression $C = ab^t$ on GDC: **(M1)**

$a = 25.5\ldots \approx 25.5, \quad b = 0.716\ldots \approx 0.716$ **A1**

Note: Accept $a \in [25.2, 25.8]$, $b \in [0.714, 0.718]$.

(b) $R^2 = 0.999\ldots \approx 0.999$ **A1**

(c) $C(8) = 25.5 \times (0.716)^8 = 25.5 \times 0.0687\ldots = 1.75\ldots \approx 1.75$ mg L^{-1} **A1**

This is an extrapolation beyond the data range ($t > 5$), so the prediction may be unreliable as the model may not hold for larger values of t. **R1**

Note: FT from their a and b.

Q12 — Back-Transform from Linearised Regression [MEDIUM]

(a) Intercept: $\log_{10} a = 0.699 \Rightarrow a = 10^{0.699} = 5.00\ldots \approx 5.00$ g **A1**

Gradient: $\log_{10} b = 0.082 \Rightarrow b = 10^{0.082} = 1.208\ldots \approx 1.21$ **A1**

(b) $M(12) = 5.00 \times (1.208)^{12} = 5.00 \times 8.916\ldots = 44.6\ldots \approx 44.6$ g **A1**

Note: FT from their a and b.

(c) Any valid assumption, e.g. "the exponential growth rate remains constant", or "the relationship $M = ab^w$ holds beyond the range of data collected." **R1**

Q13 — Power Regression: Pendulum [MEDIUM]

(a) A log-log transformation is appropriate because $\log_{10} T = \log_{10} a + b \log_{10} L$ is linear in $\log_{10} L$, so the points on a log-log graph lie on a straight line if the power model is correct. **R1**

(b) Power regression on GDC: **(M1)**

$a = 0.200\ldots \approx 0.200, \quad b = 0.502\ldots \approx 0.502$ **A1**

Note: Accept $a \in [0.198, 0.202]$, $b \in [0.500, 0.504]$.

(c) $R^2 = 0.9999\ldots$ (essentially 1.000). **A1**

The power model is highly appropriate: R^2 is extremely close to 1, indicating the model explains virtually all the variation in T. **R1**

Q14 — Decibel Inversion and Ratio [MEDIUM]

(a) $94 = 10\log_{10}\left(\dfrac{I_P}{10^{-12}}\right)$ **M1**

$\log_{10}\left(\dfrac{I_P}{10^{-12}}\right) = 9.4 \Rightarrow I_P = 10^{9.4-12} = 10^{-2.6} = 2.512\ldots \times 10^{-3}$ W m^{-2} **A1**

$I_P = 2.51 \times 10^{-3}$ W m^{-2} **A1**

(b) $I_{\text{after}} = 10^{7.9-12} = 10^{-4.1}$ W m^{-2}

$\dfrac{I_{\text{before}}}{I_{\text{after}}} = \dfrac{10^{-2.6}}{10^{-4.1}} = 10^{1.5} = 31.6\ldots \approx 31.6$ **M1 A1**

Note: Accept $\dfrac{I_B}{I_A} = 10^{(94-79)/10} = 10^{1.5}$.

Q15 — Model Selection: Quadratic vs Logarithmic [MEDIUM]

(a) Quadratic regression on GDC: **(M1)**

$R = -14.4d^2 + 126d - 26.4$ (coefficients to 3 s.f.) **A1**

Note: Accept small rounding differences. Award A1 for all three coefficients correct to 3 s.f.

(b) The quadratic model is more appropriate. **A1**

$R^2 = 0.998 > 0.969$, so the quadratic model explains a higher proportion of the variance in daily revenue.
R1

(c) $R(10) = -14.4(100) + 126(10) - 26.4 = -1440 + 1260 - 26.4 = -206$ USD. **(A1)**

This is negative (physically meaningless), so the model is unreliable for $d = 10$: it is an extrapolation well beyond the data range and the quadratic eventually decreases. **R1**

Note: Award the R1 for any valid reliability comment; the negative prediction itself earns the A mark.

Q16 — Semi-Log Graph: Back-Transform [MEDIUM]

(a) Gradient $= \dfrac{2.10 - 0.60}{5 - 0} = 0.30$; intercept $= 0.60$.

$\log_{10} y = 0.30x + 0.60$ **A1**

(b) $\log_{10} a = 0.60 \Rightarrow a = 10^{0.60} = 3.981\ldots \approx 3.98$ **A1**

$\log_{10} b = 0.30 \Rightarrow b = 10^{0.30} = 1.995\ldots \approx 2.00$ **A1**

(c) $y = 3.98 \times (2.00)^3 = 3.98 \times 8 = 31.8\ldots \approx 31.8$ **A1**

Note: FT from their a and b.

Q17 — Power Regression via Log-Log Transform [MEDIUM]

(a) Transformed data table:

$\log_{10} h$	0.602	0.903	1.079	1.204
$\log_{10} v$	0.785	0.863	0.903	0.934

A1

(b) Linear regression of $\log_{10} v$ on $\log_{10} h$: **(M1)**

Gradient $= b = 0.246\ldots \approx 0.246$; intercept $= \log_{10} a = 0.636\ldots$ **A1**

$a = 10^{0.636\ldots} = 4.33\ldots \approx 4.33$ **A1**

So $v = 4.33\, h^{0.246}$

Note: Accept $a \in [4.30, 4.36]$, $b \in [0.244, 0.248]$.

Q18 — Log Algebra and Exact Solve [MEDIUM]

(a) $f(x) = \log_{10}((x+3)(x-1)) = \log_{10}(x^2+2x-3)$ **A1**

(b) $\log_{10}(x^2+2x-3) = 1 \Rightarrow x^2+2x-3 = 10$ **M1**

$x^2+2x-13 = 0 \Rightarrow x = \dfrac{-2 \pm \sqrt{4+52}}{2} = \dfrac{-2 \pm \sqrt{56}}{2} = -1 \pm \sqrt{14}$ **A1**

$x = -1+\sqrt{14} \approx 2.742$ (domain $x > 1$: accepted); $\quad x = -1-\sqrt{14} \approx -4.742$ (rejected, $x \not> 1$) **A1**

$x = -1+\sqrt{14}$ **A1**

Note: Award A1 for correct quadratic, A1 for exact roots, A1 for discarding invalid root.

Q19 — Power Regression: Metabolic Scaling [MEDIUM]

(a) Power regression $B = am^b$ on GDC: **(M1)**

$a = 3.45\ldots \approx 3.45, \quad b = 0.752\ldots \approx 0.752$ **A1**

Note: Accept $a \in [3.40, 3.50]$, $b \in [0.748, 0.756]$.

(b) $R^2 = 0.9997\ldots \approx 0.9997$

Approximately 99.97% of the variation in metabolic rate is explained by the power model, indicating an excellent fit. **R1**

(c) Since $b \approx 0.75 < 1$, metabolic rate increases at a sub-linear rate with body mass: larger mammals have lower metabolic rates *per unit mass* than smaller ones. **R1**

Q20 — Exponential Regression and Residual [MEDIUM]

(a) Exponential regression $N = ab^t$ on GDC: **(M1)**

$a = 2.00\ldots \approx 2.00, \quad b = 1.78\ldots \approx 1.78; \quad R^2 = 0.9999\ldots \approx 1.000$ **A1**

Note: Accept $a \in [1.98, 2.02]$, $b \in [1.77, 1.79]$.

(b) Predicted value at $t = 2$: $N = 2.00 \times (1.78)^2 = 2.00 \times 3.168\ldots = 6.336\ldots$ **(M1)**

Residual $= 6.3 - 6.34\ldots = -0.034\ldots \approx -0.034$ (millions) **A1**

A negative residual means the model *over-predicts* the observed value at $t = 2$: the actual count is slightly below the fitted value. **R1**

Note: FT from their a and b; accept residual in $[-0.05, 0.05]$.

Q21 — Exponential Model + Linearisation + Prediction [HARD]

(a) GDC exponential regression $P = ab^t$: **(M1)**

$a = 33.0\ldots \approx 33.0, b = 1.0527\ldots \approx 1.05; \quad R^2 = 0.9998\ldots \approx 1.000$ **A1**

Note: Accept $a \in [32.8, 33.2]$, $b \in [1.052, 1.054]$.

(b) $\log_{10} P = t \log_{10} b + \log_{10} a$; linear regression of $\log_{10} P$ on t: **(M1)**

Gradient $= \log_{10} b = 0.02230\ldots \Rightarrow b = 10^{0.02230\ldots} = 1.0527\ldots \approx 1.053$

Intercept $= \log_{10} a = 1.519\ldots \Rightarrow a = 10^{1.519\ldots} = 33.1\ldots \approx 33.1$ **A1**

Note: Award A1 for both a and b correct to 3 s.f. from the linearised route.

(c) At $t = 25$: $P = 33.1 \times (1.053)^{25} = 33.1 \times 3.673\ldots = 121.6\ldots > 120$. **A1**

Yes, the model predicts the population will exceed 120 thousand within 25 years.

Q22 — Log-Log: Gradient, Power Model, Unknown k [HARD]

(a) On the log-log axes, gradient $= \dfrac{\log_{10} 720 - \log_{10} 45}{\log_{10} 12 - \log_{10} 3}$ **M1**

$= \dfrac{\log_{10} 16}{\log_{10} 4} = \dfrac{4 \log_{10} 2}{2 \log_{10} 2} = 2$ **A1**

(AG)

(b) $b = 2$. Using $(3, 45)$: $45 = a \cdot 3^2 \Rightarrow a = 5$. **M1**

$y = 5x^2; \quad a = 5, b = 2$ **A1**

(c) $5000 = 5k^2 \Rightarrow k^2 = 1000 \Rightarrow k = \sqrt{1000} = 31.6\ldots \approx 31.6$ **A1**

Q23 — Decibels: Combined Intensity [HARD]

(a) $72 = 10\log_{10}(I_P/10^{-12}) \Rightarrow \log_{10}(I_P) = 7.2 - 12 = -4.8$ **M1**

$I_P = 10^{-4.8} = 1.585\ldots \times 10^{-5} \approx 1.58 \times 10^{-5}$ W m^{-2} **A1**

(b) $L_Q = 72 + 15 = 87$ dB **(M1)**

$I_Q = 10^{(87-120)/10} = 10^{-3.3} = 5.012\ldots \times 10^{-4} \approx 5.01 \times 10^{-4}$ W m^{-2} **A1**

(c) $I_{\text{total}} = I_P + I_Q = 1.585 \times 10^{-5} + 5.012 \times 10^{-4} = 5.171 \times 10^{-4}$ W m^{-2} **M1**

$L_{\text{total}} = 10\log_{10}\left(\dfrac{5.171 \times 10^{-4}}{10^{-12}}\right) = 10\log_{10}(5.171 \times 10^{8}) = 87.1\ldots \approx 87.1$ dB **A1**

Q24 — Linearisation Show-That, Find b, Criterion [HARD]

(a) $D = 5 \cdot b^t \Rightarrow \log_{10} D = \log_{10} 5 + t\log_{10} b.$ **M1**

This is linear in t with gradient $\log_{10} b$ and intercept $\log_{10} 5$. **A1**

(AG)

(b) Compute $\log_{10} D$ for each t: $\log_{10} D = 0.969, 1.258, 1.540, 1.827, 2.109.$

Linear regression of $\log_{10} D$ on t: gradient $= 0.2850\ldots \approx 0.285$ **(M1)**

$\log_{10} b = 0.285 \Rightarrow b = 10^{0.285} = 1.929\ldots \approx 1.93$ **A1**

Note: Accept $b \in [1.92, 1.94]$.

(c) At $t = 8$: $D = 5 \times (1.93)^8 = 5 \times 53.1\ldots = 265\ldots < 500.$ **(M1)**

Check $t = 9$: $D = 5 \times (1.93)^9 = 5 \times 102\ldots = 513\ldots > 500.$

The criterion is **not** met: D first exceeds 500 thousand at $t = 9$ months, not $t = 8$. **A1**

Note: FT from their b.

Q25 — Exponential Fit vs Linearised Regression [HARD]

(a) GDC ExpReg $R = a \cdot b^{\theta}$: **(M1)**

$a = 10.6\ldots \approx 10.6, b = 1.00502\ldots \approx 1.005; \quad R^2 = 0.9996\ldots \approx 1.000$ **A1**

(b) Compute $Y = \log_{10} R$: $Y = 1.1072, 1.1644, 1.2175, 1.2718, 1.3222, 1.3747.$

Linear regression of Y on θ: **(M1)**

Gradient $= \log_{10} b = 0.002176\ldots \Rightarrow b = 10^{0.002176\ldots} = 1.00502\ldots \approx 1.005$ **A1**

Intercept $= \log_{10} a = 1.0635\ldots \Rightarrow a = 10^{1.0635\ldots} = 11.57\ldots \approx 11.6$ **A1**

(c) The GDC exponential regression minimises the sum of squared residuals on the *original* scale of R; the linearised regression minimises the sum of squared residuals on the *log* scale ($\log_{10} R$). These are different optimisation criteria, so the fitted parameters differ slightly. **R1**

Q26 — Log-Log: Unknown k, Exact a, Third Point [HARD]

(a) On the log-log graph, gradient $= \dfrac{\log_{10} k - \log_{10} 8}{\log_{10} 16 - \log_{10} 4} = \dfrac{\log_{10} k - \log_{10} 8}{\log_{10} 4}$ **M1**

Since the gradient equals b: $b = \dfrac{\log_{10} k - \log_{10} 8}{\log_{10} 4}$. **A1**

(AG)

(b) $1.5 = \dfrac{\log_{10} k - \log_{10} 8}{\log_{10} 4}$ **M1**

$\log_{10} k = 1.5 \log_{10} 4 + \log_{10} 8 = \log_{10} 4^{1.5} + \log_{10} 8 = \log_{10} 8 + \log_{10} 8 = \log_{10} 64$

$k = 64.$ **A1**

Using $(4, 8)$: $8 = a \cdot 4^{1.5} = a \cdot 8 \Rightarrow a = 1$. (Exact: $a = 1$.) **A1**

(c) $m = 1 \cdot 64^{1.5} = (64)^{3/2} = (8^2)^{3/2} = 8^3 = 512$ **A1**

Q27 — Braking Distance: Model Selection with Physics [HARD]

(a) GDC regressions: **(M1)**

Power model $d = av^b$: $a = 0.00266\ldots$, $b = 2.00\ldots$; $R^2 = 0.9999\ldots \approx 1.000$ **A1**

Quadratic model $d = pv^2 + qv + r$: $R^2 = 0.9998\ldots \approx 1.000$ **A1**

(b) Both models give $R^2 \approx 1.000$, so R^2 alone does not distinguish them. The power model with $b \approx 2$ is chosen: physics predicts $d \propto v^2$, and the regression confirms $b = 2.00$, making the power model both statistically and physically justified. **R1R1**

(c) $d = 0.00266 \times (130)^2 = 0.00266 \times 16900 = 44.95\ldots \approx 45.0$ m
Wait: let us recheck with the exact regression value. Using power regression: $d \approx 0.00267 \times 130^2 \approx 45.2$ m. **A1**

Reason for unreliability: extrapolation beyond the data range; road conditions, tyre friction, or driver reaction may differ at higher speeds. **R1**

Q28 — Richter + Exponential Decay Model [HARD]

(a) $M(3) = 5.8 \times (0.78)^3 = 5.8 \times 0.474552 = 2.752\ldots \approx 2.75$ **A1**

(b) Solve $5.8 \times (0.78)^n < 2.0$: **M1**

$(0.78)^n < \dfrac{2.0}{5.8} = 0.3448\ldots$

$n \ln(0.78) < \ln(0.3448\ldots) \Rightarrow n > \dfrac{\ln(0.3448\ldots)}{\ln(0.78)}$ (inequality reverses as $\ln(0.78) < 0$)

$n > \dfrac{-1.0647\ldots}{-0.2485\ldots} = 4.283\ldots$ **A1**

Since n is a positive integer, the magnitude first falls below 2.0 at $n = 5$. **A1**

(c) $\frac{I_1}{I_0} = 10^{M_1}$ and $\frac{I_5}{I_0} = 10^{M_5}$, so $\frac{I_1}{I_5} = 10^{M_1 - M_5}$ **M1**

$M_1 = 5.8 \times 0.78 = 4.524; \quad M_5 = 5.8 \times (0.78)^5 = 5.8 \times 0.2887\ldots = 1.674\ldots$ **(A1)**

$\frac{I_1}{I_5} = 10^{4.524-1.674} = 10^{2.850} = 707.9\ldots \approx 708$ **A1**

Q29 — Power Model: Show-That, Exact c and k, Predict [HARD]

(a) $\log_{10} P_1 = \log_{10} k + c \log_{10} s_1; \quad \log_{10} P_2 = \log_{10} k + c \log_{10} s_2.$

Subtracting: $c = \frac{\log_{10} P_2 - \log_{10} P_1}{\log_{10} s_2 - \log_{10} s_1} = \frac{\log_{10} 25600 - \log_{10} 3200}{\log_{10} 40 - \log_{10} 10}$ **M1**

$= \frac{\log_{10} 8}{\log_{10} 4}$ **A1**

(AG)

(b) $c = \frac{\log_{10} 8}{\log_{10} 4} = \frac{3 \log_{10} 2}{2 \log_{10} 2} = \frac{3}{2}$ **A1**

(c) Using $(10,\ 3200)$: $3200 = k \cdot 10^{3/2} \Rightarrow k = \frac{3200}{10\sqrt{10}} = \frac{320}{\sqrt{10}} = 32\sqrt{10}$ **A1**

(d) $P(25) = 32\sqrt{10} \cdot 25^{3/2} = 32\sqrt{10} \cdot 125 = 4000\sqrt{10} = 12649\ldots \approx \$12\,600$ **A1**

Q30 — Species-Area Law: Full Log-Log Analysis [HARD]

(a) Compute $\log_{10} A$: $0, 0.602, 1.204, 1.806, 2.408, 3.010.$

Compute $\log_{10} S$: $1.079, 1.301, 1.519, 1.740, 1.959, 2.182.$

Linear regression of $\log_{10} S$ on $\log_{10} A$: **(M1)**

Gradient $= z = 0.3660\ldots \approx 0.366$; intercept $= \log_{10} c = 1.079\ldots \Rightarrow c = 10^{1.079\ldots} = 12.0\ldots \approx 12.0$ **A1**

Model: $S = 12.0\, A^{0.366}$

Note: Accept $c \in [11.8, 12.2]$, $z \in [0.364, 0.368]$.

(b) $R^2 = 0.9999\ldots \approx 1.000.$

The model is extremely suitable: virtually all variation in species count is explained by island area via the power law. **R1**

(c) $S(100) = 12.0 \times 100^{0.366} = 12.0 \times 5.848\ldots = 70.2\ldots > 40.$ **(M1)**

The model predicts approximately 70 species, which **supports** the ecologist's claim. **A1**

(d) Any valid limitation, e.g. "for very large islands, species diversity may plateau due to habitat saturation or competition effects that the power law does not model", or "the model is an extrapolation for islands larger than those in the data." **R1**

Common Mistakes to Avoid

1. **Confusing R^2 with r.** For non-linear models, the GDC reports R^2 directly. Do *not* square the correlation coefficient r (which is only defined for linear regression). Always use R^2 as the goodness-of-fit measure for non-linear models.

2. **Back-transforming incorrectly.** From $\log_{10} y = mx + c$, the model is $y = 10^c \cdot (10^m)^x$, so $a = 10^{\text{intercept}}$ and $b = 10^{\text{gradient}}$. A very common error is to write $b = m$ (the gradient itself) rather than $b = 10^m$.

3. **Power model vs exponential model.** A *semi-log* graph (log y vs x) indicates an exponential model $y = ab^x$. A *log-log* graph (log y vs log x) indicates a power model $y = ax^b$. Mixing these up reverses the entire analysis.

4. **Ratio calculations on logarithmic scales.** If two quantities differ by d units on the Richter scale, their intensity ratio is 10^d, **not** d times. Similarly for decibels: a 10 dB difference corresponds to a factor of 10 in intensity.

5. **Extrapolation vs interpolation.** Always comment on whether a prediction lies *within* the data range (interpolation, generally reliable) or *outside* it (extrapolation, potentially unreliable). Marks are awarded specifically for this comment.

6. **R^2 is not sufficient to compare models of different types.** R^2 compares how well a model fits the data it was fitted to, but a more complex model (e.g. cubic) will always have $R^2 \geq$ a simpler model. Use physical reasoning and parsimony as additional criteria.