PHOTON ACADEMY · IB MATH AI HL CHEATSHEET

Non-Linear Modelling & Regression

SL 2.5–2.6 · AHL 2.9–2.10

ibtuition.sg · free

FORMULAS YOU MUST KNOW

the fit minimises $\sum(y_i - \hat{y}_i)^2$ — the total squared gap between data and model

Least squares:

$y = k\,a^x$ — constant percentage growth or decay

Exponential model:

$y = a\,x^b$ — many physical and allometric laws

Power model:

R^2 close to 1 means the model explains most of the variation

Fit quality:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

Gradient:

$a = \bar{y} - b\bar{x}$

Intercept:

substitute the required x into $y = a + bx$ to get $\hat{y}$

Predict:

$-1 \le r \le 1$ — the sign gives the direction, the size gives the strength

Correlation:

$R^2 = r^2$ — the proportion of the variation in y explained by the model

Determination:

$R^2 = 0.96 \Rightarrow$ the model explains 96% of the variation in y

Reading it:

$\text{residual} = y - \hat{y}$ — observed minus predicted (from the fitted model)

Residual:

$\text{SS}_{\text{res}} = \sum(y - \hat{y})^2$

Total misfit:

KEY METHODS

- On the GDC pick the regression that matches the shape — LinReg, QuadReg, CubicReg, ExpReg or PwrReg — and it returns the parameters and R^2 in a single step.
- Choose by shape first, then confirm with R^2 and the residual pattern — never on R^2 alone. A high-order polynomial can hug the points yet behave nonsensically just outside the data.
- Find the gradient first, then the intercept, then substitute to predict. On the GDC, LinReg hands you a and b directly — the formula is only for a "show that".
- The regression line of y on x always passes through the mean point $(\bar{x}, \bar{y})$. Use it to predict y from x — not the reverse.

EXAM TRAPS TO AVOID

- Only trust predictions for x inside the data range (interpolation). Reaching outside the range is extrapolation, and the pattern may simply not continue.
- Exponential and power models look alike but differ: $y = k\,a^x$ raises a fixed base a to the variable x, whereas $y = a\,x^b$ raises the variable x to a fixed exponent b.
- Squaring destroys the sign: a strong negative $r = -0.98$ still gives a positive $R^2 = 0.960$. So $R^2 \ge 0$ always — you can never read the direction of the relationship off R^2.
- Keep the order observed − predicted: a positive residual means the point lies above the model.

- For a non-linear fit the GDC reports R^2 directly (there is no single signed r). A higher R^2 means a better fit, but a fit is only as good as its residuals — check those too (§5).
- A good model leaves residuals scattered randomly about zero. A clear curved pattern in the residuals is the data telling you the model family is wrong — switch curves.
- A power model $y = a\,x^b$ suits allometric and physical laws (e.g. period against length). Only x carries the exponent, so $(ax)^b$ is wrong; for $b > 0$ the curve passes through the origin.

Square each residual before adding, so positives and negatives cannot cancel.

- To evaluate $y = k\,a^x$, raise a to the power x first, then multiply by k — never compute $k \cdot a \cdot x$. The exponent applies to a, not to the product.
- "First reaches / first exceeds" always rounds UP (the ceiling), never to the nearest whole number: $t = 9.1$ means it first happens during period 10.
- The intercept is $\log_{10} k$ (or $\ln a$), not k (or a). Undo it with $10^{(\,)}$ for a $\log_{10}$ axis, but with $e^{(\,)}$ for a $\ln$ axis — match the base to the axis.