# Using Logs to Analyse Nonlinear Data

In stats, we develop techniques for analysing the strength of correlation and an optimal line of best fit for linear data i.e. determining an equation in the form $y = mx + c$ to model the data.

How can we similarly analyse other trends? In particular, if data seems to fit a relationship of the form $y = ab^x$ or $y = ax^b$ how can we determine suitable values of $a$ and $b$ to model the data?

To use our linear tools on these non-linear patterns, we take logs of the equations

| Form: | $y = ab^x$ | $y = ax^b$ |
|---|---|---|
| Rearrangement: | $\log(y) = \log(ab^x)$<br>$\log(y) = \log(a) + \log(b^x)$<br>$\log(y) = \log(b)\,x + \log(a)$ | $\log(y) = \log(ax^b)$<br>$\log(y) = \log(a) + \log(x^b)$<br>$\log(y) = b\log(x) + \log(a)$ |

Both have been converted to the form $y' = mx' + c$ if we choose the variables/axes correctly.
So plotting should give us a straight line (roughly) which we can analyse using our linear data tools.

| Original form: | $y = ab^x$ | $y = ax^b$ |
|---|---|---|
| Rearranged form: | $\log(y) = \log(b)\,x + \log(a)$ | $\log(y) = b\log(x) + \log(a)$ |
| $y'$-axis variable: | $\log(y)$ | $\log(y)$ |
| $x'$-axis variable: | $x$ | $\log(x)$ |
| LoBF gradient: | $\log(b)$ | $b$ |
| LoBF $y$-intercept: | $\log(a)$ | $\log(a)$ |

## The Method

- Add rows to the data table for $\log(y)$ (and $\log(x)$) if necessary
- Plot the converted data on a graph
- Plot a line of best fit
- Read the gradient and $y$-intercept off your line of best fit
- Undo the log(s) as necessary to find the values of $a$ and $b$

For the time being, you are expected to keep things simple: they will give you the equations form and expect you to hand-draw a line of best fit.

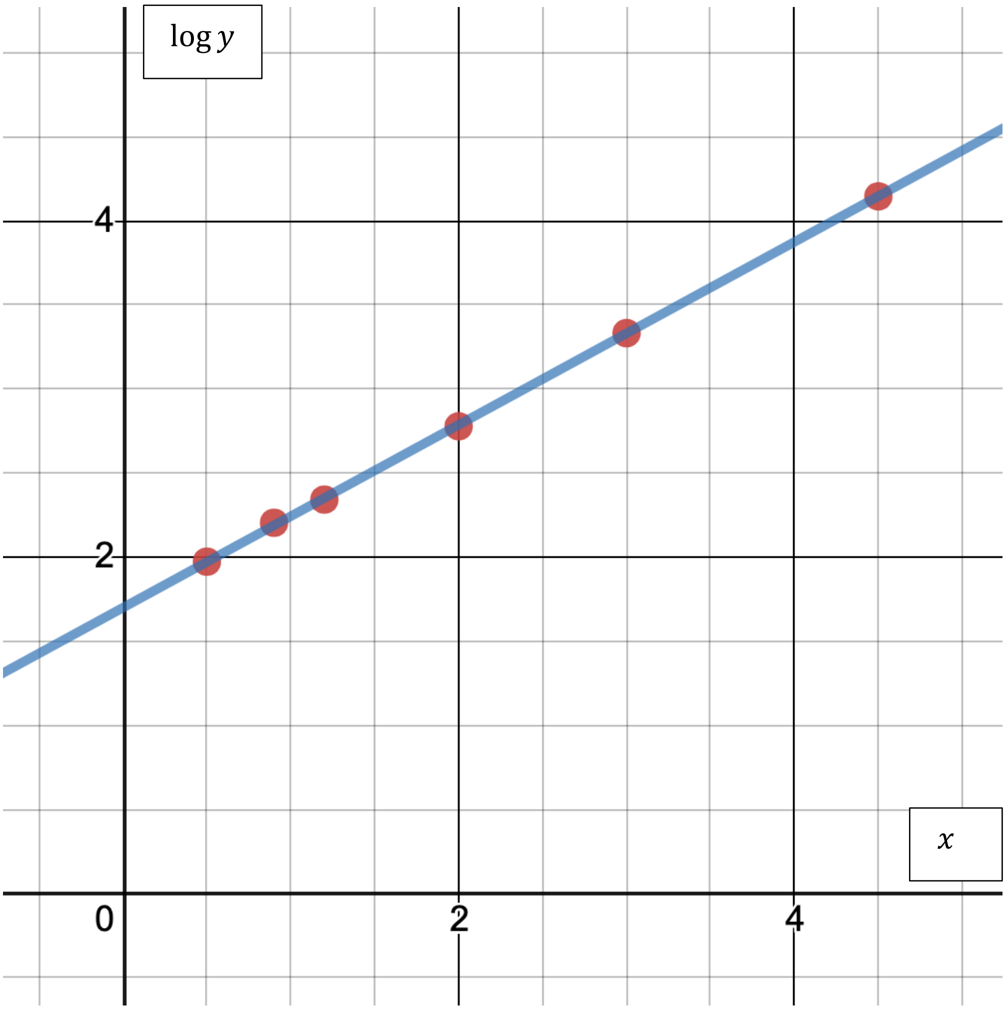
Example overleaf

e.g. Jurgen believes his data fits an exponential relationship $y = ab^x$. Plot a suitable graph and line of best fit and use them to determine approximate values for $a$ and $b$.

| $x$ | 2 | 4.5 | 0.9 | 0.5 | 3 | 1.2 |
|---|---|---|---|---|---|---|
| $y$ | 600 | 14000 | 160 | 94 | 2150 | 220 |

Add the row:

| | | | | | | |
|---|---|---|---|---|---|---|
| $\log(y)$ | 2.78 | 4.15 | 2.20 | 1.97 | 3.33 | 2.34 |




I have plotted $x$ against $\log y$ and a pretty decent line of best fit, which looks like it has a gradient of about 0.55 and $y$-intercept of about 1.7.

$$\log(b) \approx 0.55 \Rightarrow b \approx 10^{0.55} = 3.55$$
$$\log(a) \approx 1.7 \Rightarrow a \approx 10^{1.7} = 50.12$$

So the data can be modelled approximately by the relationship $y = 50.12 \times 3.55^x$