Regression, correlation and hypothesis testing 1A

1 **a** As noted at the beginning of Section 1.1, the equation $Y = 1.2 + 0.4X$ can be rewritten as $\log y = 1.2 + 0.4\log x$, which is of the form $\log y = \log a + n\log x$ and so $y = ax^n$.

b $Y = 1.2 + 0.4X$

$\Rightarrow \log y = 1.2 + 0.4\log x$

$\Rightarrow y = 10^{1.2+0.4\log x} = 10^{1.2} \times 10^{0.4\log x}$

$\Rightarrow y = 10^{1.2} \times 10^{\log x^{0.4}} = 10^{1.2} \times x^{0.4}$

Therefore $a = 10^{1.2} \approx 15.8$ (3 s.f.) and $n = 0.4$

2 **a** As noted at the beginning of Section 1.1, the equation $Y = 0.4 + 1.6X$ can be rewritten as $\log y = 0.4 + 1.6x$, which is of the form $\log y = \log k + x\log b$ and so $y = kb^x$.

b $Y = 0.4 + 1.6X$

$\Rightarrow \log y = 0.4 + 1.6x$

$\Rightarrow y = 10^{0.4+1.6x} = 10^{0.4} \times 10^{1.6x}$

$\Rightarrow y = 10^{0.4} \times \left(10^{1.6}\right)^x$

Therefore $k = 10^{0.4} \approx 2.51$ (3 s.f.) and $b = 10^{1.6} \approx 39.8$.

3 In the linear model $Y = mX + c$, where m and c are constants,
$Y = \log y$ and $X = \log x$, so $\log y = m\log x + c$

Therefore $c = \log a$

The point (0, 172) lies on the line, so $c = 172$ and $\log a = 172 \Rightarrow a = 10^{172}$

(23, 109) lies on $Y = mX + 172$:

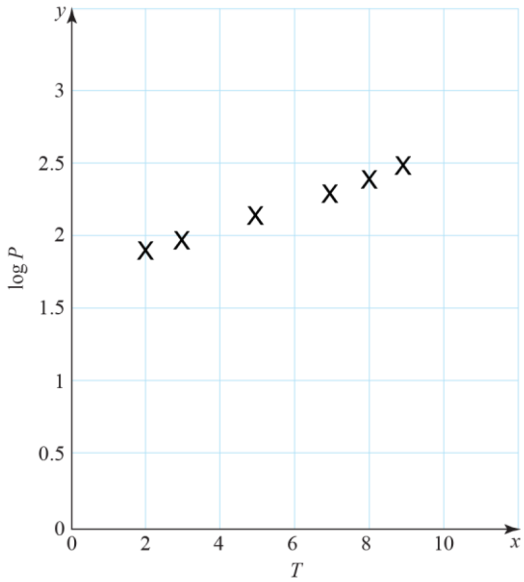
$109 = 23m + 172$

$\Rightarrow 23m = 109 - 172$

$\Rightarrow m = \frac{-63}{23} \approx -2.739$ (3 d.p.).

4 a

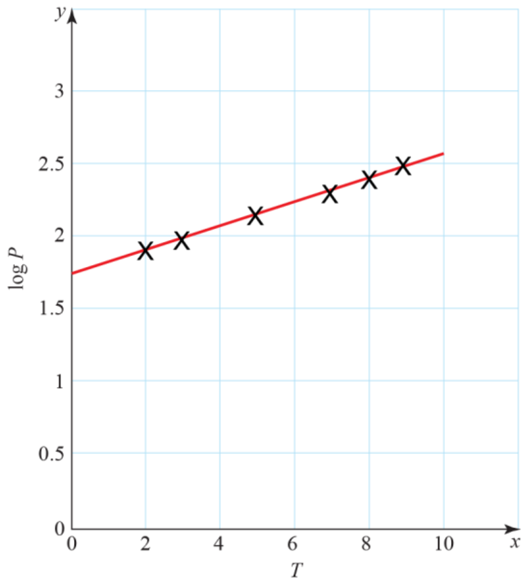
T	2	3	5	7	8	9
log *P*	1.86	1.93	2.10	2.25	2.33	2.41



b The points seem to lie on a straight line with a positive gradient, which suggests a strong positive correlation.

c Yes – the variables show a linear relationship when log P is plotted against T.

d



If $\log P = mT + c$ then $P = 10^c(10^m)^T$. Measuring the gradient and intercept from the line of best fit with computer provides $c = 1.69927$ and $m = 0.07901$. These then give $a = 50.0345502089$ and $b = 1.19949930315$. Allow c between 1.65 and 1.8 so that a can be between 44.7 and 63.1. Since the gradient is small, it is better found using the original data points. Allow m between 0.077 and 0.082 so that b can be between 1.19 and 1.21.

4 e The approximate model is $P = 50.1 \times 1.2^T$ and so increasing T by 1 gives

$50.1 \times 1.2^{T+1} = (50.1 \times 1.2^T) \times 1.2$

which means increasing T by 1 corresponds to an increase of the population by 20%. Note that T is recorded in months, and so for every month that passes, the population of moles increases by 20%.

5 a The equation $t = a + bn$ is the equation of a straight line, but the data on the scatter diagram are not close to a straight line.

b $y = -0.301 + 0.6x$

$\Rightarrow \log t = -0.301 + 0.6 \log n$

$\Rightarrow t = 10^{-0.301 + 0.6 \log n} = 10^{-0.301} \times 10^{0.6 \log n}$

$\Rightarrow t = 10^{-0.301} \times 10^{\log n^{0.6}}$

$\Rightarrow t = 10^{-0.301} \times n^{0.6}$

Therefore $a = 10^{-0.301} \approx 0.5$ (3 s.f.) and $k = 0.6$.

6 $y = 1.31x - 0.41$

$\Rightarrow \log r = 1.31 \log c - 0.41$

$\Rightarrow r = 10^{1.31 \log c - 0.41} = 10^{1.31 \log c} \times 10^{-0.41}$

$\Rightarrow r = 10^{\log c^{1.31}} \times 10^{-0.41} = c^{1.31} \times 10^{-0.41}$

Therefore $r = 0.389 \times c^{1.31}$ (3 s.f.).

7 $y = 0.0023 + 1.8x$

$\Rightarrow \log m = 0.0023 + 1.8 \log h$

$\Rightarrow m = 10^{0.0023 + 1.8 \log h} = 10^{0.0023} \times 10^{1.8 \log h}$

$\Rightarrow m = 10^{0.0023} \times 10^{\log h^{1.8}} = 10^{0.0023} \times h^{1.8}$

Therefore $a = 10^{0.0023} \approx 1.0$ (3 s.f.) and $n = 1.8$.

8 a $y = 0.09 + 0.05x$

$\Rightarrow \log g = 0.09 + 0.05t$

$\Rightarrow g = 10^{0.09 + 0.05t} = 10^{0.09} \times 10^{0.05t}$

$\Rightarrow g = 10^{0.09} \times \left(10^{0.05}\right)^t$

Therefore $a = 10^{0.09} \approx 1.23$ and $b \approx 1.12$ (3 s.f.)

b If you increase the temperature by 1 °C, b is the increase in the growth rate g, i.e. b is the rate of change of g per degree.

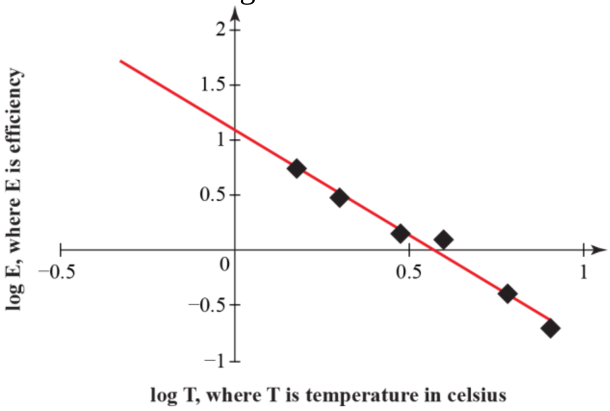
c 35 °C is outside of the range of data (extrapolation).

Challenge

a Construct a table of values for log$T$ and log$E$:

log *T*	0.0792	0.176	0.301	0.477	0.602	0.778	0.903
log *E*	0.954	0.74	0.477	0.146	0.0969	−0.398	−0.699

Plot the scatter diagram and draw a line of best fit:



The fact that the data are fitted by a straight line shows the validity of the relationship.

b The y-intercept of the line of best fit is 1.1 (to 2 s.f.).
So log a = 1.1 (approximately)
$a = 10^{1.1} = 12.58925\ldots = 12.6$
From the graph, the gradient of the line of best fit is approximately −1.90 (to 3 s.f.), so b = −1.90.

c The model is of the form $\log E = \log a + b\log T$, but the expression $\log a + b\log T$ is not defined when $T = 0$ since $\log(0)$ is undefined (it approaches $-\infty$ as $T \to 0^+$).