Stats2 Chapter 1: Measuring Correlation

Exponential Models

Chapter Overview

1:: Exponential Models

Recap of Pure Year 1. Using $y = ab^x$ to model an exponential relationship between two variables.

2:: Measuring Correlation

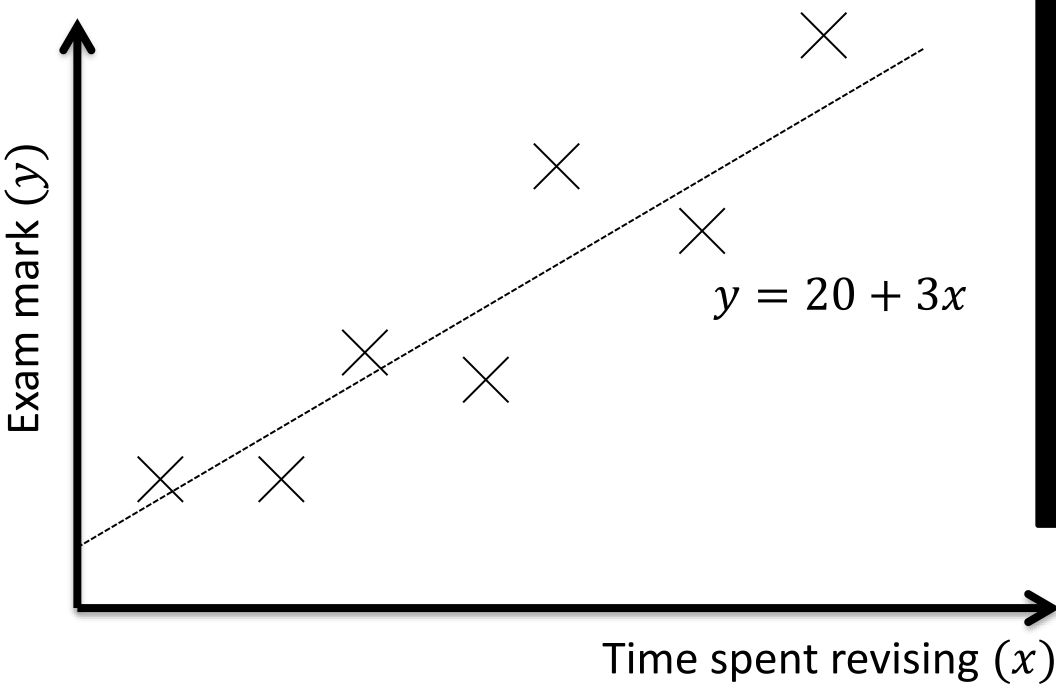
Using the Product Moment Correlation Coefficient (PMCC), r, to measure the strength of correlation between two variables.

3:: Hypothesis Testing for no correlation

We want to test whether two variables have some kind of correlation, or whether any correlation observed just happened by chance.

Teacher Notes: (1) is mostly a recap of Pure Year 1. (2) is in the old S1 module, but students now just use their calculator to calculate r; they do not need to use formulae. (3) is from the old S3 module but simplified.

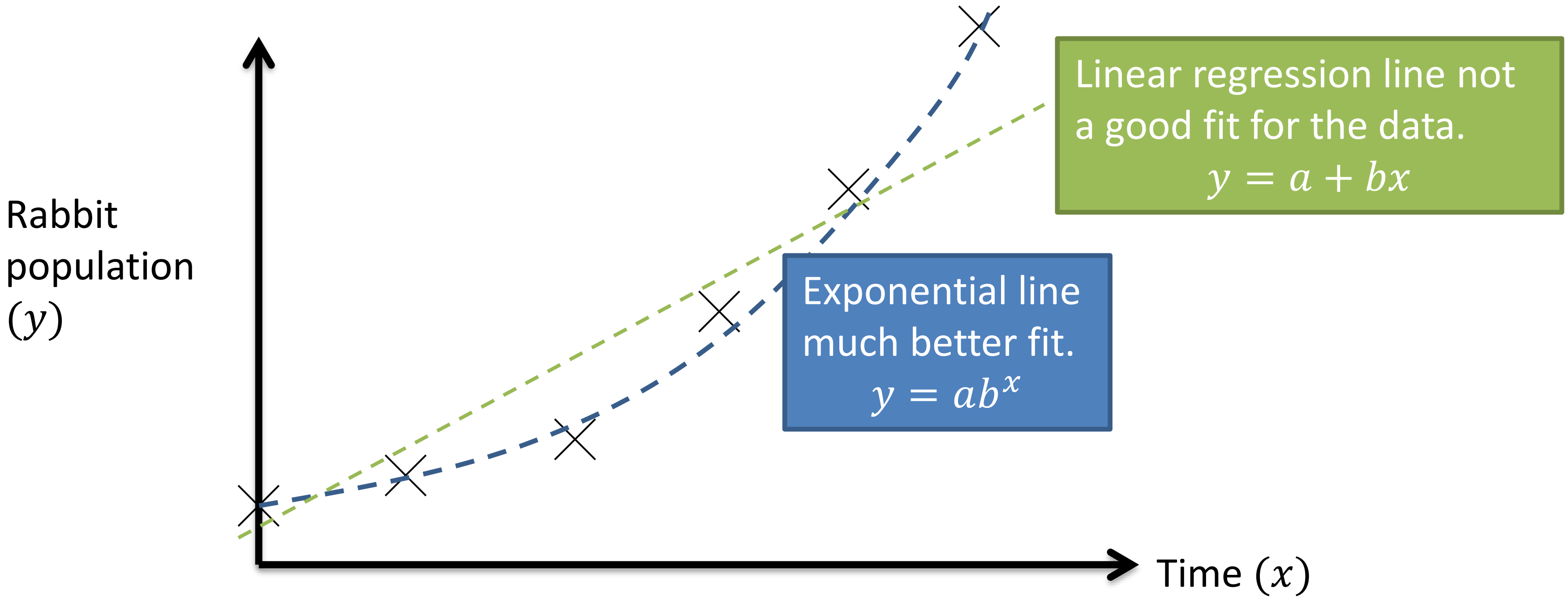
RECAP :: What is regression?



What we've done here is come up with a **model** to explain the data, in this case, a line $y = a + bx$. We've then tried to set a and b such that the resulting y value matches the actual exam marks as closely as possible.
The 'regression' bit is the act of setting the parameters of our model (here the gradient and y-intercept of the line of best fit) to best explain the data.

I record people's exam marks as well as the time they spent revising. I want to predict how well someone will do based on the time they spent revising. How would I do this?

Exponential Regression



For some variables, e.g. population with time, it may be more appropriate to use an **exponential** equation, i.e. $y = ab^x$, where a and b are constants we need to fix to best match the data.

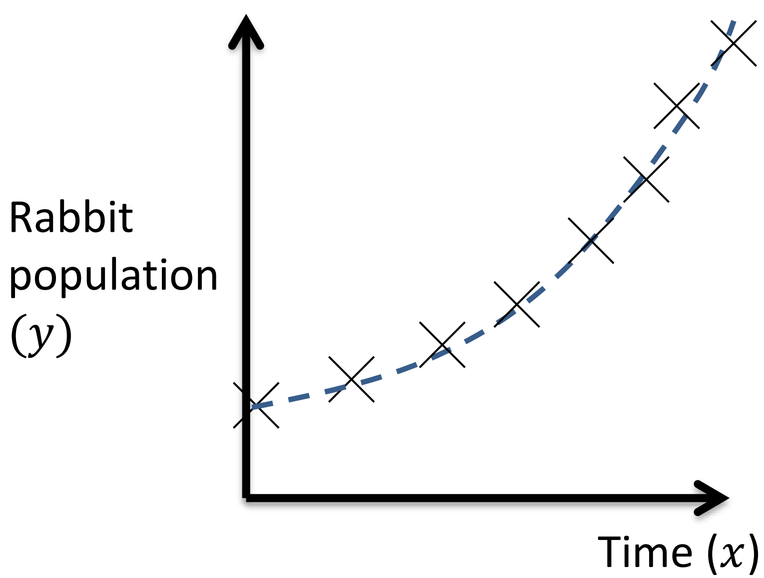
$$y = ab^x$$
$$\log y = \log(ab^x)$$
$$\log y = \log a + x \log b$$

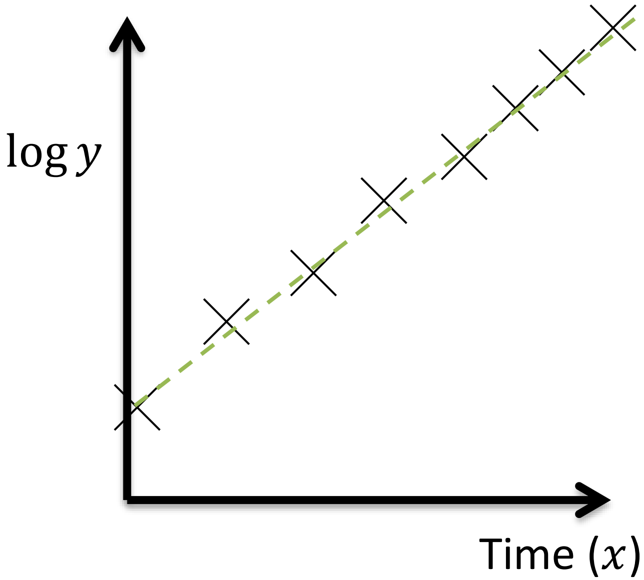
In Year 1, what did we do to both sides to end up with a straight line equation?

✎ If $y = kb^x$ for constants k and b then $\log y = \log k + x \log b$

Exponential Regression

If $y = kb^x$ for constants k and b then $\log y = \log k + x \log b$





Comparing the equations, we can see that if we log the y values (although leave the x values), the data then forms a straight line, with y-intercept $\log k$ and gradient $\log b$.

Example

[Textbook] The table shows some data collected on the temperature, in °C, of a colony of bacteria (t) and its growth rate (g).

Temperature, t (°C)	3	5	6	8	9	11
Growth rate, g	1.04	1.49	1.79	2.58	3.1	4.46

The data are coded using the changes of variable $x = t$ and $y = \log g$. The regression line of y on x is found to be $y = -0.2215 + 0.0792x$.
a. Mika says that the constant -0.2215 in the regression line means that the colony is shrinking when the temperature is 0°C. Explain why Mika is wrong
b. Given that the data can be modelled by an equation of the form $g = kb^t$ where k and b are constants, find the values of k and b.

a

?

b

?

Example

[Textbook] The table shows some data collected on the temperature, in °C, of a colony of bacteria (t) and its growth rate (g).

Temperature, t (°C)	3	5	6	8	9	11
Growth rate, g	1.04	1.49	1.79	2.58	3.1	4.46

The data are coded using the changes of variable $x = t$ and $y = \log g$. The regression line of y on x is found to be $y = -0.2215 + 0.0792x$.
a. Mika says that the constant -0.2215 in the regression line means that the colony is shrinking when the temperature is 0°C. Explain why Mika is wrong
b. Given that the data can be modelled by an equation of the form $g = kb^t$ where k and b are constants, find the values of k and b.

a When $t = 0, y = -0.2215 + (0.0792 \times 0) = -0.2215$

$$y = \log g \ \therefore g = 10^y = 10^{-0.2215} = 0.600\ (3sf)$$

The growth rate is positive.

b $g = kb^t$

$\log g = \log k + t \log b$

Compare to $y = -0.2215 + 0.0792x \ \rightarrow \ \log g = -0.2215 + 0.0792g$

$$\log k = -0.2215 \ \rightarrow \ k = 10^{-0.2215} = 0.600$$
$$\log b = 0.0792 \ \rightarrow \ b = 10^{0.0792} = 1.20$$

The textbooks starts with $y = -0.2215 + 0.0792x$ and raises 10 to the power of each side. Alternatively start with $g = kb^t$ and then log.

Test Your Understanding

Robert wants to model a rabbit population P with respect to time in years t. He proposes that the population can be modelled using an exponential model: $P = kb^t$ The data is coded using $x = t$ and $y = \log P$. The regression line of y on x is found to be $y = 2 + 0.3x$. Determine the values of k and b.

?

Rabbit

MyHomework.com has had the following annual numbers of page views each year:

Year	2012	2013	2014	2015	2016	2017	2018	2019
Views V	10115	26790	60306	180386	1119801	8.3 m	21.9 m	*57.5 m*

projected

Determine the appropriate constants if we assume a polynomial model $V = at^b$, where t is the number of years after 2011.

?

Test Your Understanding

Robert wants to model a rabbit population P with respect to time in years t. He proposes that the population can be modelled using an exponential model: $P = kb^t$ The data is coded using $x = t$ and $y = \log P$. The regression line of y on x is found to be $y = 2 + 0.3x$. Determine the values of k and b.

$$P = kb^t$$
$$\log P = \log k + t \log b$$
$$y = \log k + x \log b$$
$$\therefore \log k = 2 \quad \rightarrow \quad k = 100$$
$$\log b = 0.3 \quad \rightarrow \quad b = 2.00\ (3sf)$$

Rabbit

MyHomework.com has had the following annual numbers of page views each year:

Year	2012	2013	2014	2015	2016	2017	2018	2019
Views V	10115	26790	60306	180386	1119801	8.3 m	21.9 m	*57.5 m*

projected

Determine the appropriate constants if we assume a polynomial model $V = at^b$, where t is the number of years after 2011.

$$V = at^b$$
$$\log V = \log a + b \log t$$

So need to log the t values and V values.

log t	0	0.3010	0.4771	0.6020	0.6989	0.7781	0.8450	0.9030
log V	4.0050	4.4280	4.7804	5.2562	6.0491	6.9191	7.3404	7.7597

Using calculator stats mode, y-intercept: $\mathbf{\log a = 3.3406}$, gradient: $\mathbf{b = 4.3025}$

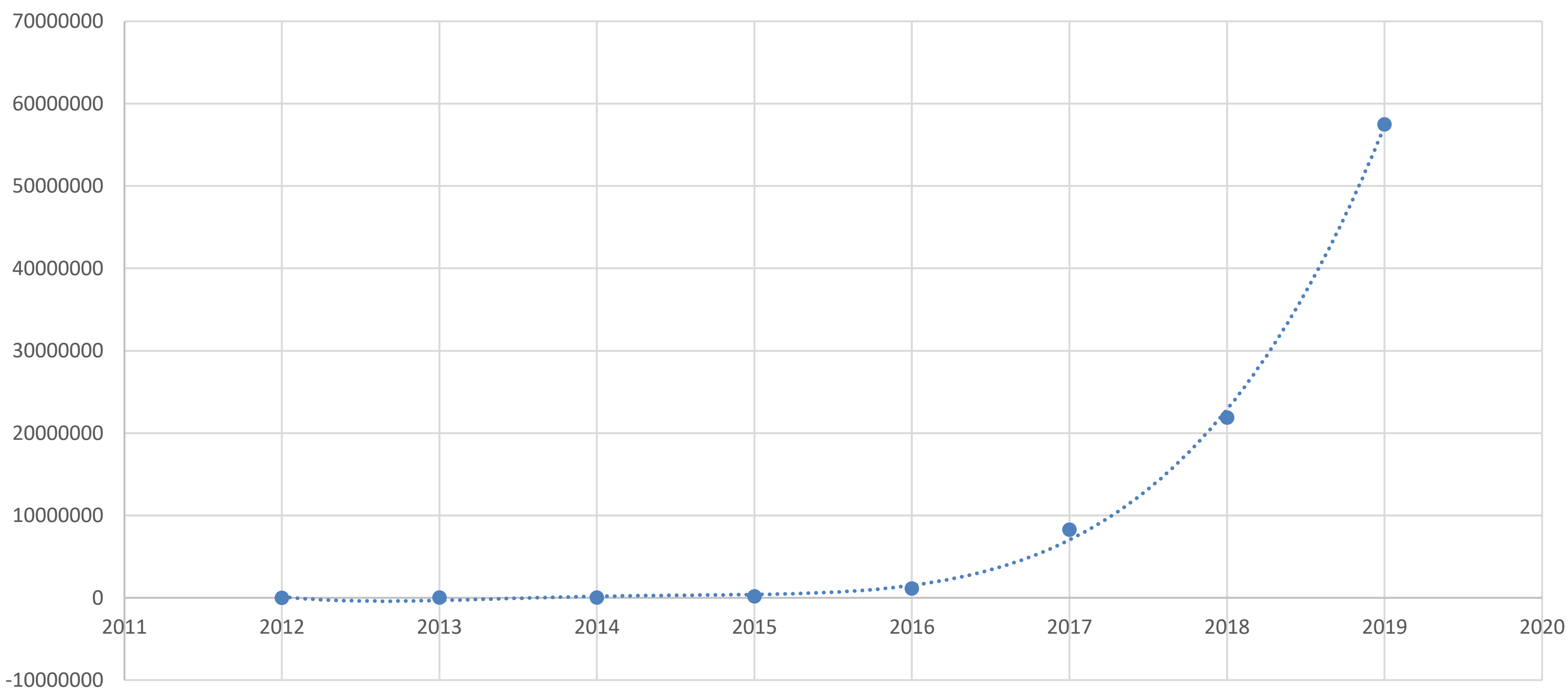
Therefore $\mathbf{V = 2191 \times t^{4.3025}}$

DrFrostMaths has had the following annual numbers of page views each year:

Year	2012	2013	2014	2015	2016	2017	2018	2019
Views V	10115	26790	60306	180386	1119801	8.3 m	21.9 m	57.5 m

Determine the appropriate constants if we assume a polynomial model $V = at^b$, where t is the number of years after 2011.

As you can see below, this polynomial model we found produces a strong fit to the data! An exponential model would have also produced a good fit.



Exercise 1.1

Pearson Stats/Mechanics Year 2

Pages 1-2

Homework Exercise

1 Data are coded using $Y = \log y$ and $X = \log x$ to give a linear relationship. The equation of the regression line for the coded data is $Y = 1.2 + 0.4X$.

a State whether the relationship between y and x is of the form $y = ax^n$ or $y = kb^x$.

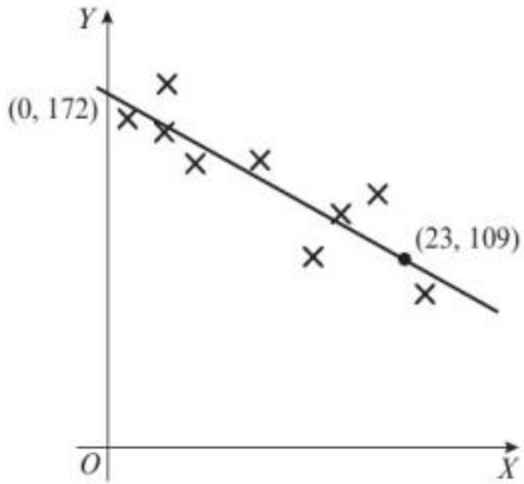
b Write down the relationship between y and x and find the values of the constants.

2 Data are coded using $Y = \log y$ and $X = x$ to give a linear relationship. The equation of the regression line for the coded data is $Y = 0.4 + 1.6X$.

a State whether the relationship between y and x is of the form $y = ax^n$ or $y = kb^x$.

b Write down the relationship between y and x and find the values of the constants.

3 The scatter diagram shows the relationship between two sets of coded data, X and Y, where $X = \log x$ and $Y = \log y$. The regression line of Y on X is shown, and passes through the points (0, 172) and (23, 109).



The relationship between the original data sets is modelled by an equation of the form $y = ax^n$. Find, correct to 3 decimal places, the values of a and n.

Homework Exercise

4 The size of a population of moles is recorded and the data are shown in the table. T is the time, in months, elapsed since the beginning of the study and P is the number of moles in the population.

T	2	3	5	7	8	9
P	72	86	125	179	214	257

a Plot a scatter diagram showing $\log P$ against T.

b Comment on the correlation between $\log P$ and T.

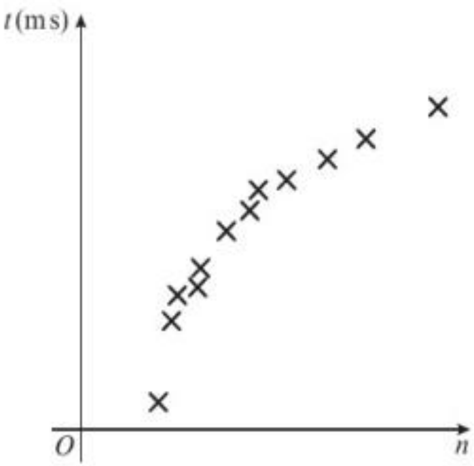
c State whether your answer to **b** supports the fact that the original data can be modelled by a relationship of the form $P = ab^T$.

d Approximate the values of a and b for this model.

e Give an interpretation of the value of b you calculated in part **d**.

Hint Think about what happens when the value of T increases by 1. When interpreting coefficients, refer in your answer to the context given in the question.

5 The time, t m s, needed for a computer algorithm to determine whether a number, n, is prime is recorded for different values of n. A scatter graph of t against n is drawn.



a Explain why a model of the form $t = a + bn$ is unlikely to fit these data.

The data are coded using the changes of variable $y = \log t$ and $x = \log n$. The regression line of y on x is found to be $y = -0.301 + 0.6x$.

b Find an equation for t in terms of n, giving your answer in the form $t = an^k$, where a and k are constants to be found.

Homework Exercise

7 The heights, h cm, and masses, m kg, of a sample of Galapagos penguins are recorded. The data are coded using $y = \log m$ and $x = \log h$ and it is found that a linear relationship exists between x and y. The equation of the regression line of y on x is $y = 0.0023 + 1.8x$.

Find an equation to describe the relationship between m and h, giving your answer in the form $m = ah^n$, where a and n are constants to be found.

8 The table shows some data collected on the temperature, t °C, of a colony of insect larvae and the growth rate, g, of the population.

Temp, t (°C)	13	17	21	25	26	28
Growth rate, g	5.37	8.44	13.29	20.91	23.42	29.38

The data are coded using the changes of variable $x = t$ and $y = \log g$. The regression line of y on x is found to be $y = 0.09 + 0.05x$.

a Given that the data can be modelled by an equation of the form $g = ab^t$ where a and b are constants, find the values of a and b. **(3 marks)**

b Give an interpretation of the constant b in this equation. **(1 mark)**

c Explain why this model is not reliable for estimating the growth rate of the population when the temperature is 35 °C. **(1 mark)**

Challenge

The table shows some data collected on the efficiency rating, E, of a new type of super-cooled engine when operating at a certain temperature, T.

Temp, T (°C)	1.2	1.5	2	3	4	6	8
Efficiency, E	9	5.5	3	1.4	0.8	0.4	0.2

It is thought that the relationship between E and t is of the form $E = aT^b$.

a By plotting an appropriate scatter diagram, verify that this relationship is valid for the data given.

b By drawing a suitable line on your scatter diagram and finding its equation, estimate the values of a and b.

c Give a reason why the model will not predict the efficiency of the engine when the temperature is 0 °C.

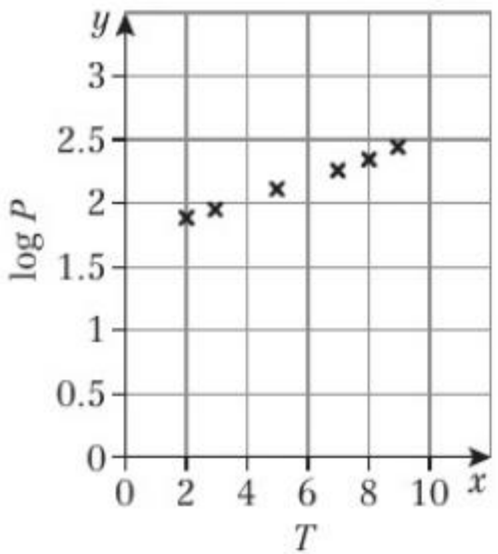
Homework Answers

1 **a** $y = ax^n$ **b** $a = 15.8$ (3 s.f.), $n = 0.4$

2 **a** $y = kb^x$ **b** $k = 2.51$, $b = 39.8$ (3 s.f.)

3 $a = 1 \times 10^{172}$, $n = -2.739$ (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 Strong positive correlation

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

d $a = 50.1$ (3 s.f.), $b = 1.2$

e For every month that passes, the population of moles increases by 20%.

5 **a** $t = a + bn$ would show a linear relationship. This graph is not a straight line.

b $a = 0.5$, $k = 0.6$

6 $r = 0.389c^{1.31}$

7 $a = 1.0$, $n = 1.8$

8 **a** $a = 1.23$ (3 s.f.), $b = 1.12$ (3 s.f.)

b b is the rate of change of g per degree.

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

Challenge

a A graph of log T against log E shows a straight line.

b $\log E = 1.09 - 1.96(\log T)$, $a = 12.3$ (3 s.f.), $b = -1.96$ (3 s.f.)

c log 0 is undefined.