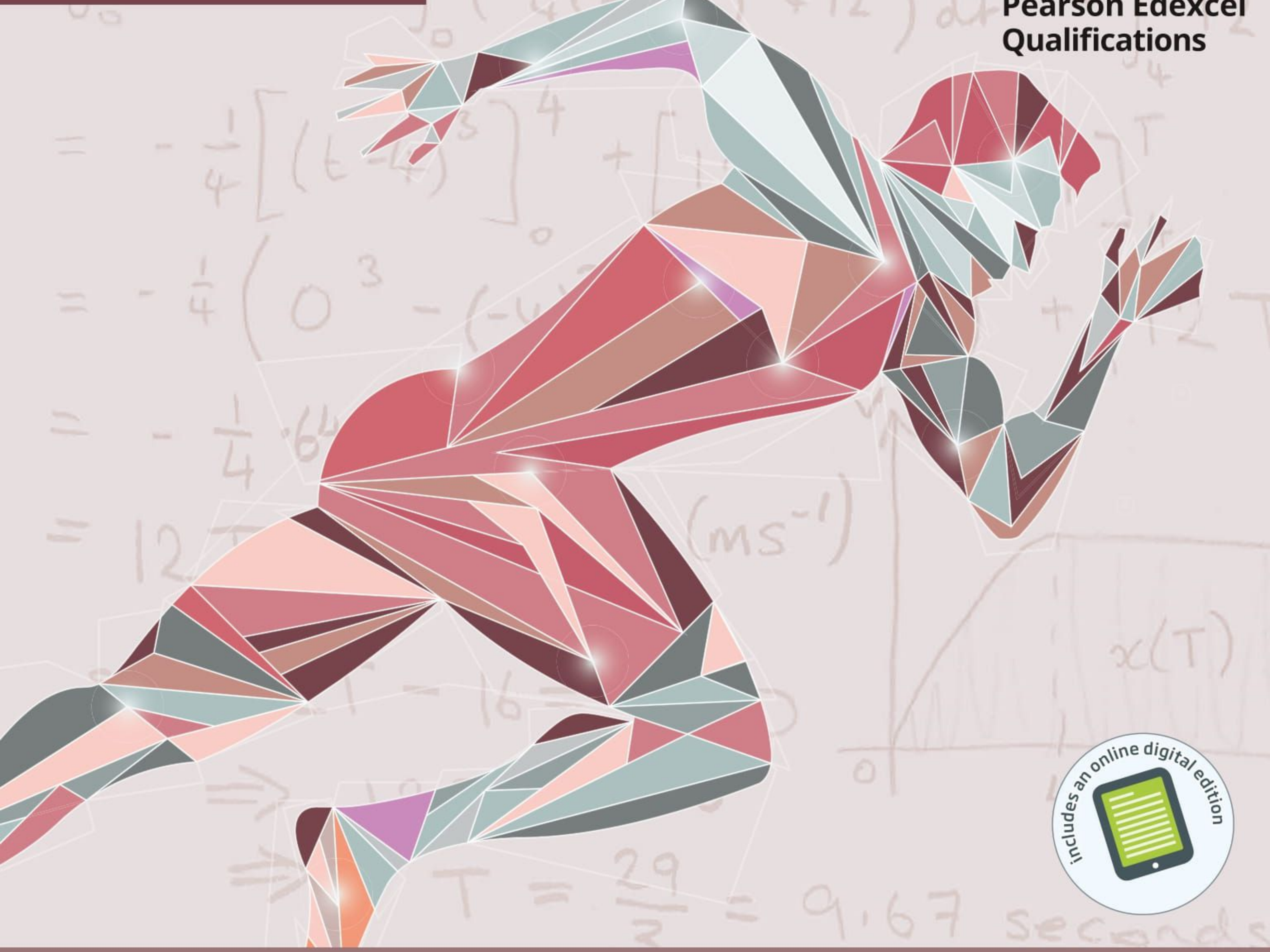
11 - 19 PROGRESSION

Endorsed for
Pearson Edexcel Qualifications



includes an online digital edition

Pearson Edexcel A level Mathematics

Statistics and Mechanics

Year 2

Pearson

Pearson Edexcel A level Mathematics

Statistics and Mechanics Year 2

Series Editor: Harry Smith

Pearson's market-leading books are the most trusted resources for Pearson Edexcel AS and A level Mathematics.

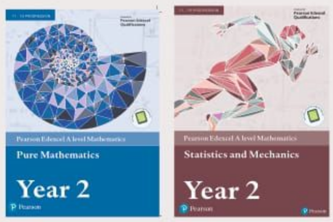
This book can be used alongside the Year 1 book to cover all the content needed for the **Pearson Edexcel A level Statistics and Mechanics exam**.

- Fully updated to match the 2017 specifications, with more of a focus on problem-solving and modelling as well as supporting the large data set and new calculators.
- FREE additional online content to support your independent learning, including full worked solutions for every question in the book (SolutionBank), GeoGebra interactives and Casio calculator tutorials.
- Includes access to an online digital edition (valid for 3 years once activated).
- Includes worked examples with guidance, lots of exam-style questions, a practice paper, and plenty of mixed and review exercises.

Pearson Edexcel AS and A level Mathematics books

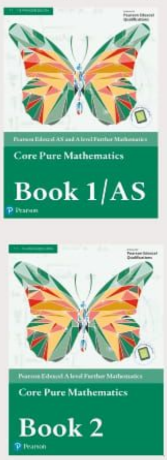
Year 1/AS

Year 2



Pearson Edexcel AS and A level Further Mathematics books

Compulsory



Options

For more information visit: www.pearsonschools.co.uk/edalevelmaths2017

www.pearsonschools.co.uk
myorders@pearson.com

ISBN 978-1-4469-4407-3

9 781446 944073

Pearson

Contents

Overarching themes

The following three overarching themes have been fully integrated throughout the Pearson Edexcel AS and A level Mathematics series, so they can be applied alongside your learning and practice.

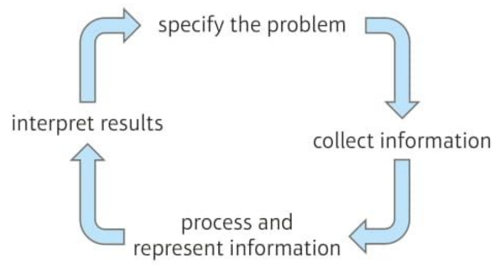
1. Mathematical argument, language and proof

- Rigorous and consistent approach throughout
- Notation boxes explain key mathematical language and symbols
- Dedicated sections on mathematical proof explain key principles and strategies
- Opportunities to critique arguments and justify methods

2. Mathematical problem solving

- Hundreds of problem-solving questions, fully integrated into the main exercises
- Problem-solving boxes provide tips and strategies
- Structured and unstructured questions to build confidence
- Challenge boxes provide extra stretch

The Mathematical Problem-solving cycle

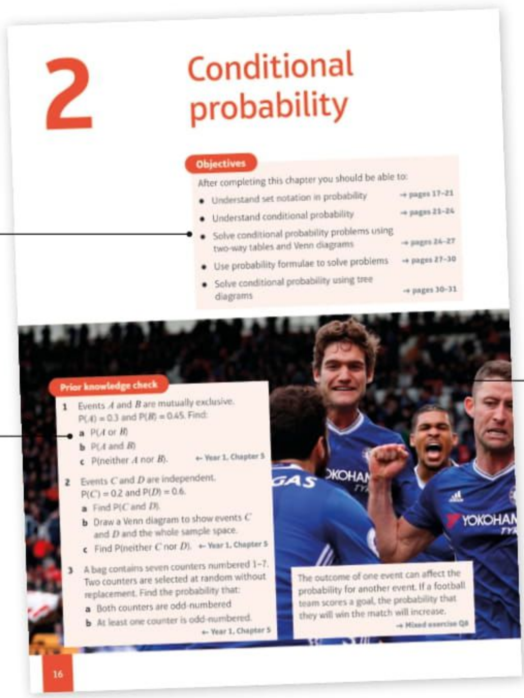


3. Mathematical modelling

- Dedicated modelling sections in relevant topics provide plenty of practice where you need it
- Examples and exercises include qualitative questions that allow you to interpret answers in the context of the model
- Dedicated chapter in Statistics & Mechanics Year 1/AS explains the principles of modelling in mechanics

Finding your way around the book

Access an online digital edition using the code at the front of the book.



Each chapter starts with a list of objectives

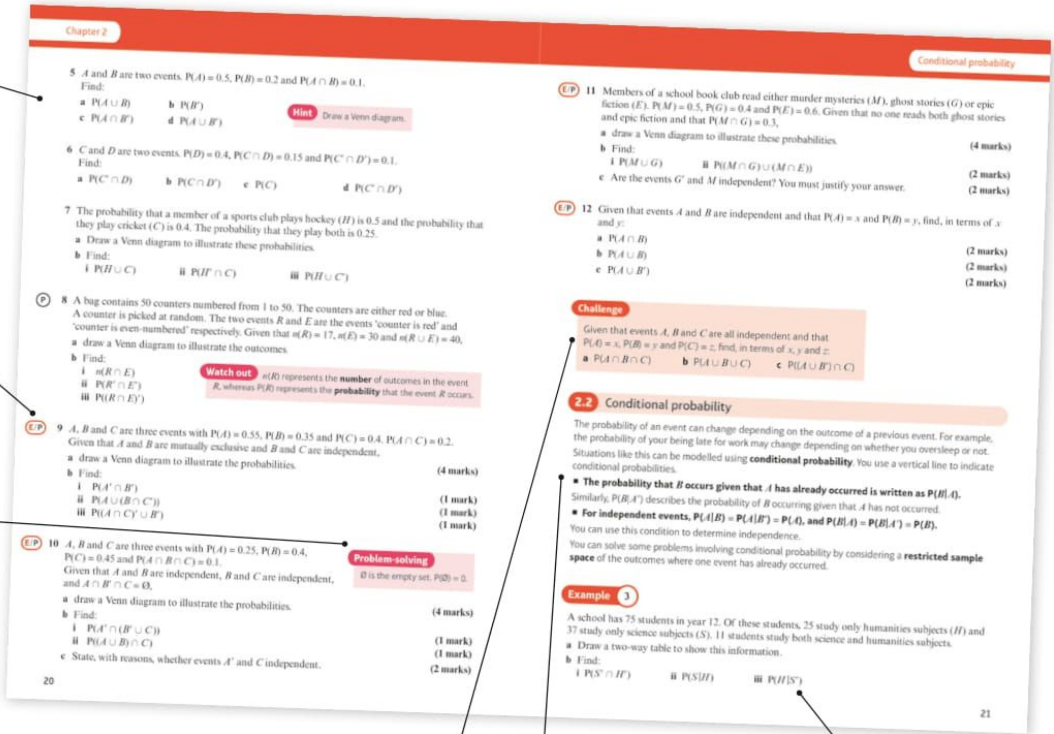
The *Prior knowledge check* helps make sure you are ready to start the chapter

The real world applications of the maths you are about to learn are highlighted at the start of the chapter with links to relevant questions in the chapter

Exercise questions are carefully graded so they increase in difficulty and gradually bring you up to exam standard

Exercises are packed with exam-style questions to ensure you are ready for the exams

Problem-solving boxes provide hints, tips and strategies, and *Watch out* boxes highlight areas where students often lose marks in their exams



Exam-style questions are flagged with (E)

Problem-solving questions are flagged with (P)

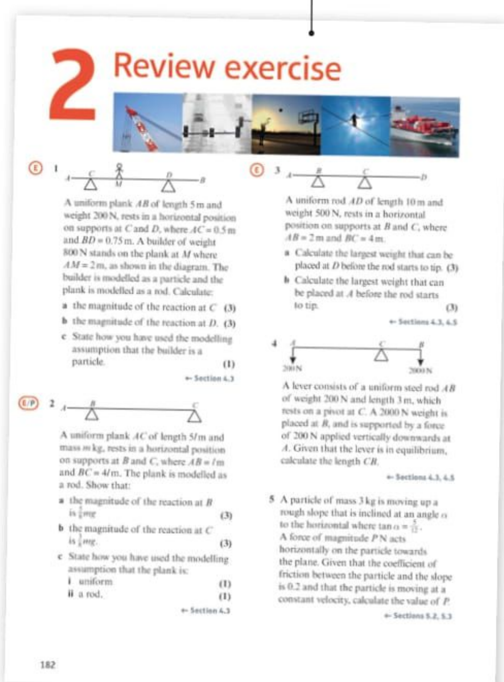
Each chapter ends with a *Mixed exercise* and a *Summary of key points*

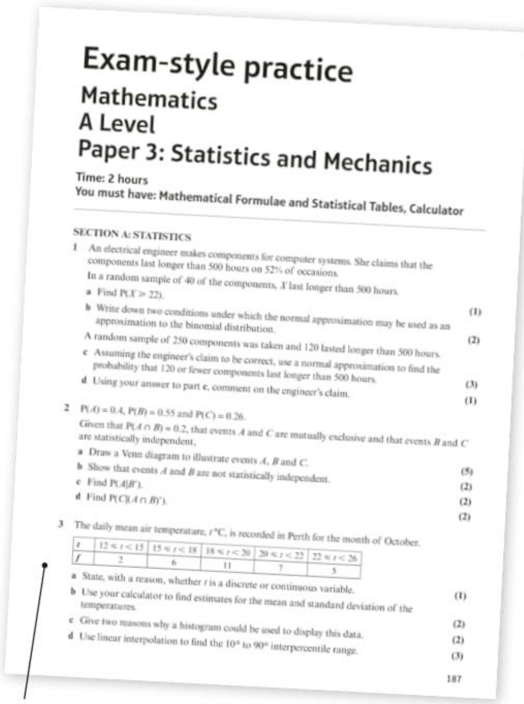
Challenge boxes give you a chance to tackle some more difficult questions

Each section begins with explanation and key learning points

Step-by-step worked examples focus on the key types of questions you'll need to tackle

Every few chapters a *Review exercise* helps you consolidate your learning with lots of exam-style questions





Two A level practice papers at the back of the book help you prepare for the real thing

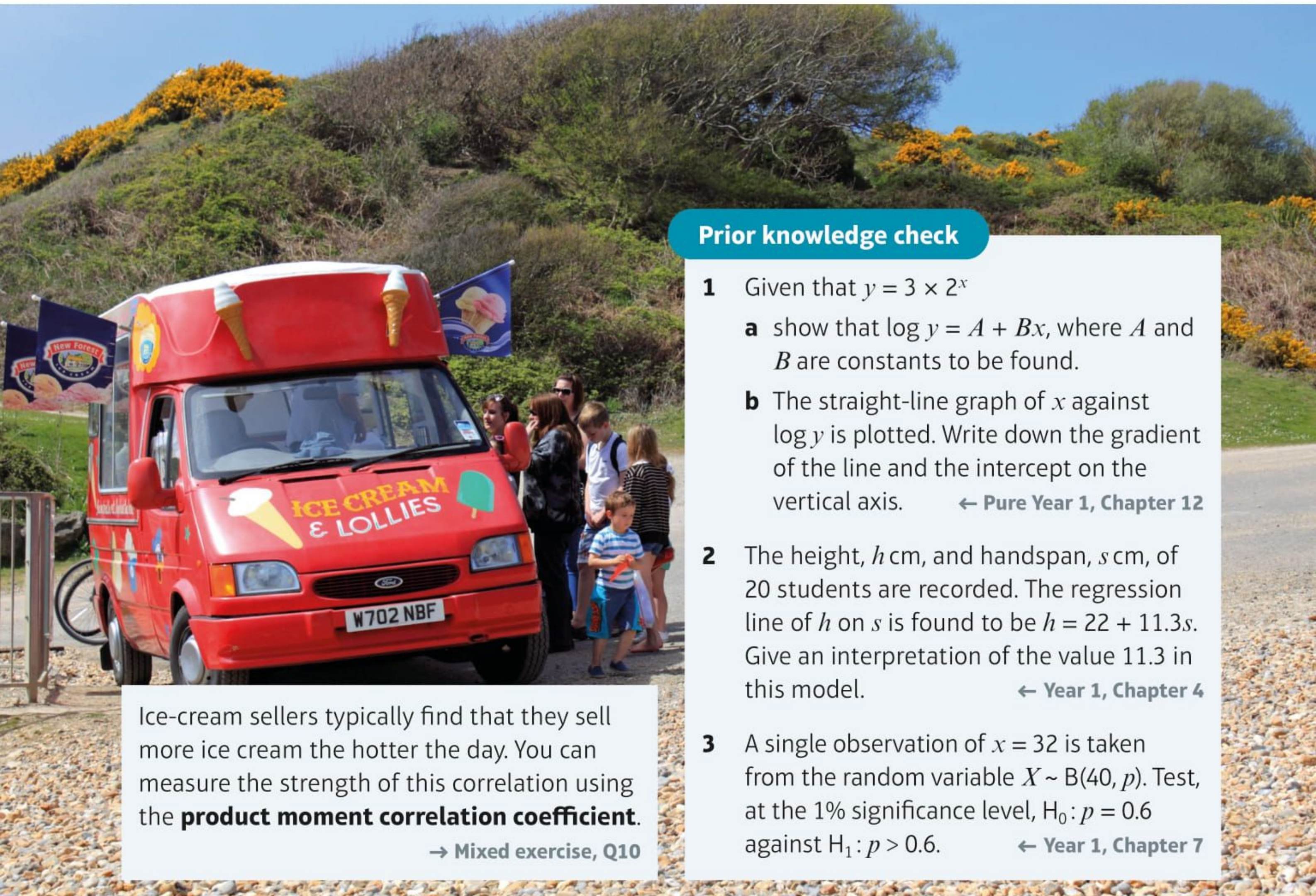
Regression, correlation and hypothesis testing

1

Objectives

After completing this chapter you should be able to:

- Understand exponential models in bivariate data → pages 2–5
- Use a change of variable to estimate coefficients in an exponential model → pages 2–5
- Understand and calculate the product moment correlation coefficient → pages 5–8
- Carry out a hypothesis test for zero correlation → pages 8–12



Ice-cream sellers typically find that they sell more ice cream the hotter the day. You can measure the strength of this correlation using the **product moment correlation coefficient**.

→ Mixed exercise, Q10

Prior knowledge check

1 Given that $y = 3 \times 2^x$

a show that $\log y = A + Bx$, where A and B are constants to be found.

b The straight-line graph of x against $\log y$ is plotted. Write down the gradient of the line and the intercept on the vertical axis. ← Pure Year 1, Chapter 12

2 The height, h cm, and handspan, s cm, of 20 students are recorded. The regression line of h on s is found to be $h = 22 + 11.3s$. Give an interpretation of the value 11.3 in this model. ← Year 1, Chapter 4

3 A single observation of $x = 32$ is taken from the random variable $X \sim B(40, p)$. Test, at the 1% significance level, $H_0 : p = 0.6$ against $H_1 : p > 0.6$. ← Year 1, Chapter 7

1.1 Exponential models

Regression lines can be used to model a **linear** relationship between two variables. Sometimes, experimental data does not fit a linear model, but still shows a clear pattern. You can use logarithms and coding to examine trends in non-linear data.

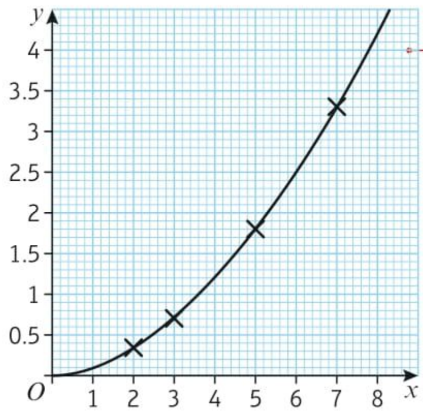
For data that can be modelled by a relationship of the form $y = ax^n$, you need to code the data using $Y = \log y$ and $X = \log x$ to obtain a linear relationship.

- **If $y = ax^n$ for constants a and n then $\log y = \log a + n \log x$**

Link Take logs of both sides and rearrange to convert the original form into the linear form.
← Pure Year 1, Section 14.8

For data that can be modelled by an **exponential** relationship of the form $y = ab^x$, you need to code the data using $Y = \log y$ and $X = x$ to obtain a linear relationship.

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

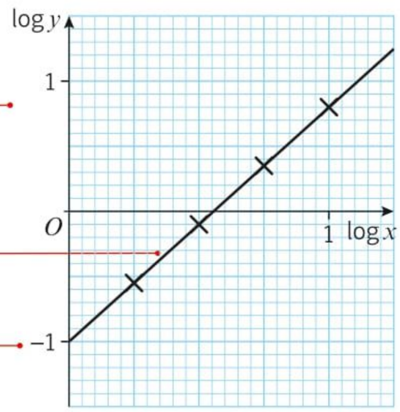


The points on this scatter graph satisfy the relationship $y = 0.1x^{1.8}$. This is in the form $y = ax^n$.

Plotting $\log x$ against $\log y$ gives a straight line.

The gradient of the line is 1.8. This corresponds to the value of n in the non-linear relationship.

The y-intercept is at (0, −1). This corresponds to $\log a$ hence $a = 10^{-1} = 0.1$, as expected.



Example 1

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 $t = 0$, $x = 0$, so according to the model,
$y = -0.2215$
$\log g = -0.2215$
$g = 10^{-0.2215} = 0.600$ (3 s.f.).
This growth rate is positive: the colony is not shrinking.

b Substitute $x = t$ and $y = \log g$:
$\log g = -0.2215 + 0.0792t$
$g = 10^{-0.2215 + 0.0792t}$
$g = 10^{-0.2215} \times (10^{0.0792})^t$
$g = 0.600 \times 1.20^t$ (both values given to 3 s.f.)
So $k = 0.600$ and $b = 1.20$

Remember that the original data have been coded. Use the coding in reverse to find the corresponding value of g. You could also observe that a prediction based on $t = 0$ would be outside the range of the data so would be an example of **extrapolation**. ← Year 1, Section 4.2

Use the change of variable to find an expression for $\log g$ in terms of t. You could also compare the equation of the regression line with $\log g = \log k + t \log b$. ← Pure Year 1, Section 14.6

Remember log means log to the base 10. So $10^{\log g} = g$.

Use the laws of indices to write the expression in the form $g = kb^t$.

Exercise 1A

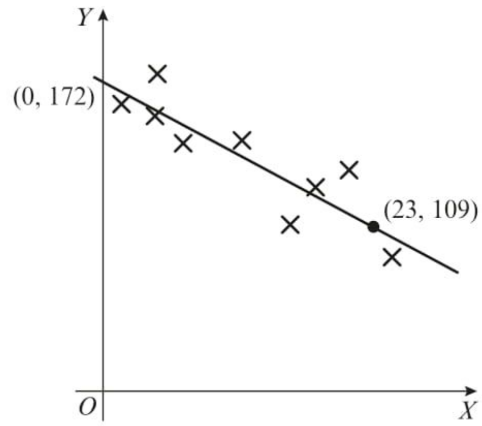
Online Explore the original and coded data graphically using technology.

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.

(P) 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 the exact value of a and the value of n correct to 3 decimal places.



(P) 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.

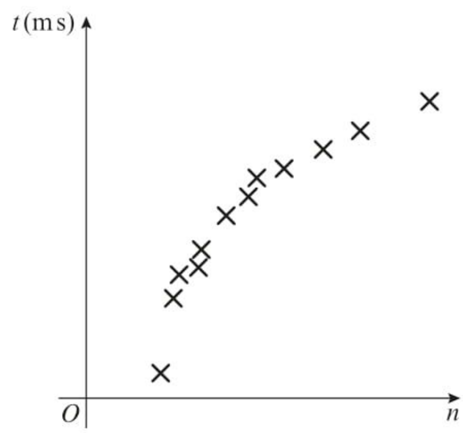
c State whether your answer to **b** supports the fact that the original data can be modelled by a relationship of the form $P = ab^T$.

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.

6 Data are collected on the number of units (c) of a catalyst added to a chemical process, and the rate of reaction (r).

The data are coded using $x = \log c$ and $y = \log r$. It is found that a linear relationship exists between x and y and that the equation of the regression line of y on x is $y = 1.31x - 0.41$.

Use this equation to determine an expression for r in terms of c.

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.

E/P **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.

1.2 Measuring correlation

You can calculate quantitative measures for the strength and type of linear correlation between two variables. One of these measures is known as the **product moment correlation coefficient**.

- **The product moment correlation coefficient describes the linear correlation between two variables. It can take values between −1 and 1.**

Notation The product moment correlation coefficient, or PMCC, for a sample of data is denoted by the letter r.

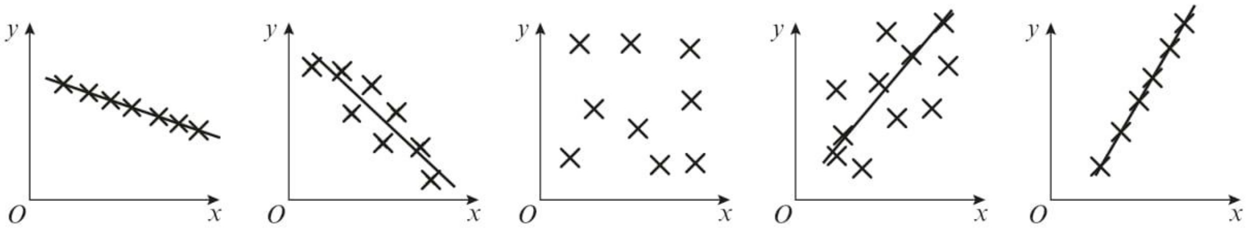
If $r = 1$, there is perfect positive linear correlation.

If $r = -1$, there is perfect negative linear correlation.

The closer r is to −1 or 1, the stronger the negative or positive correlation, respectively.

If $r = 0$ (or is close to 0) there is no linear correlation. In this case there might still be a non-linear relationship between the variables.

Hint For $r = \pm 1$, the points all lie on a straight line.



$r = -1$ $r = -0.8$ $r = 0$ $r = 0.3$ $r = 1$

You need to know how to calculate the product moment correlation coefficient for bivariate data using your calculator.

E **12** The table shows data from the large data set on the daily mean air temperature and the daily mean pressure during May and June 2015 in Beijing.

Temperature (°C)	17.5	18.5	18.0	24.6	22.2	23.1	27.3
Pressure (hPa)	1010	1011	1012	997	1009	998	1002

Test at the 2.5% level of significance the claim that there is negative correlation between the daily mean air temperature and the daily mean pressure. State your hypotheses clearly.

(4 marks)

Large data set

You will need access to the large data set and spreadsheet software to answer these questions.

1 **a** Take a random sample of size 20 from the data for Heathrow in 2015, and record the daily mean air temperature and daily total rainfall.

b Calculate the product moment correlation coefficient between these variables for your sample.

c Test, at the 5% level of significance, the claim that there is a correlation between the daily mean air temperature and the daily total rainfall.

2 **a** State with a reason whether you would expect to find a relationship between daily mean total cloud cover and daily mean visibility.

b Use a random sample from the large data set to test for this relationship. You should state clearly:

- Your sample size and location
- Your sampling method
- The hypotheses and significance level for your test
- A conclusion in the context of the question

Hint You might be able to use the **Correl** or **CorrelationCoefficient** commands in your spreadsheet software to calculate the PMCC.

Summary of key points

1 If $y = ax^n$ for constants a and n then $\log y = \log a + n \log x$

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

3 The **product moment correlation coefficient** describes the linear correlation between two variables. It can take values between −1 and 1.

4 For a one-tailed test use either:

- $H_0: \rho = 0$, $H_1: \rho > 0$ or
- $H_0: \rho = 0$, $H_1: \rho < 0$

For a two-tailed test use:

- $H_0: \rho = 0$, $H_1: \rho \neq 0$