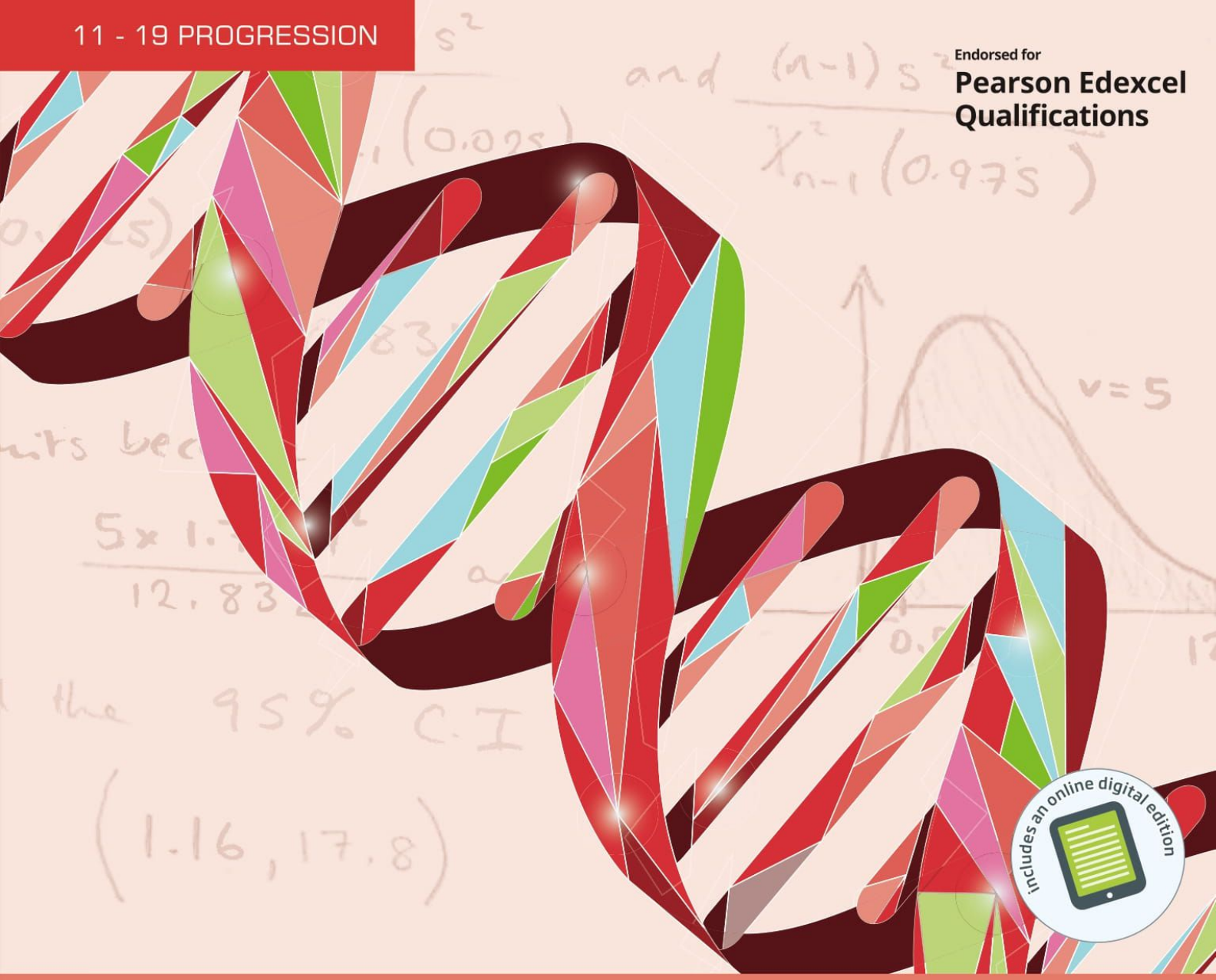
11 - 19 PROGRESSION

Endorsed for
Pearson Edexcel Qualifications



Pearson Edexcel AS and A level Further Mathematics

Further Statistics 2

FS2

Pearson

Contents

● = A level only

Overarching themes iv
Extra online content vi

1 Linear regression 1
1.1 Least squares linear regression 2
1.2 Residuals 10
Mixed exercise 1 16

2 Correlation 21
2.1 The product moment correlation coefficient 22
2.2 Spearman's rank correlation coefficient 26
2.3 Hypothesis testing for zero correlation 33
Mixed exercise 2 38

3 Continuous distributions 44
3.1 Continuous random variables 45
3.2 The cumulative distribution function 51
3.3 Mean and variance of a continuous distribution 56
3.4 Mode, median, percentiles and skewness 63
3.5 The continuous uniform distribution 71
3.6 Modelling with the continuous uniform distribution 79
Mixed exercise 3 82

● **4 Combinations of random variables** 92
● **4.1** Combinations of random variables 93

Review exercise 1 99

● **5 Estimation, confidence intervals and tests using a normal distribution** 109
● **5.1** Estimators, bias and standard error 110
● **5.2** Confidence intervals 120
● **5.3** Hypothesis testing for the difference between means 127
● **5.4** Use of large sample results for an unknown population 132
Mixed exercise 5 135

● **6 Further hypothesis tests** 141
● **6.1** Variance of a normal distribution 142
● **6.2** Hypothesis testing for the variance of a normal distribution 146
● **6.3** The F-distribution 149
● **6.4** The F-test 155
Mixed exercise 6 159

● **7 Confidence intervals and tests using the t-distribution** 163
● **7.1** Mean of a normal distribution with unknown variance 164
● **7.2** Hypothesis test for the mean of a normal distribution with unknown variance 170
● **7.3** The paired t-test 174
● **7.4** Difference between means of two independent normal distributions 180
● **7.5** Hypothesis test for the difference between means 185
Mixed exercise 7 189

● **Review exercise 2** 196

Exam-style practice paper (AS level) 204

● **Exam-style practice paper (A level)** 206

Appendix 209
Binomial cumulative distribution function 209
Percentage points of the normal distribution 214
Percentage points of the χ^2 distribution 215
Critical values for correlation coefficients 216
Percentage points of Student's t-distribution 217
Percentage points of the F-distribution 218
Poisson cumulative distribution function 219
Formulae 220

Answers 225

Index 256

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.

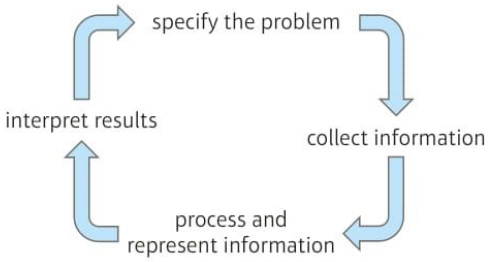
1. Mathematical argument, language and proof

- 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 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

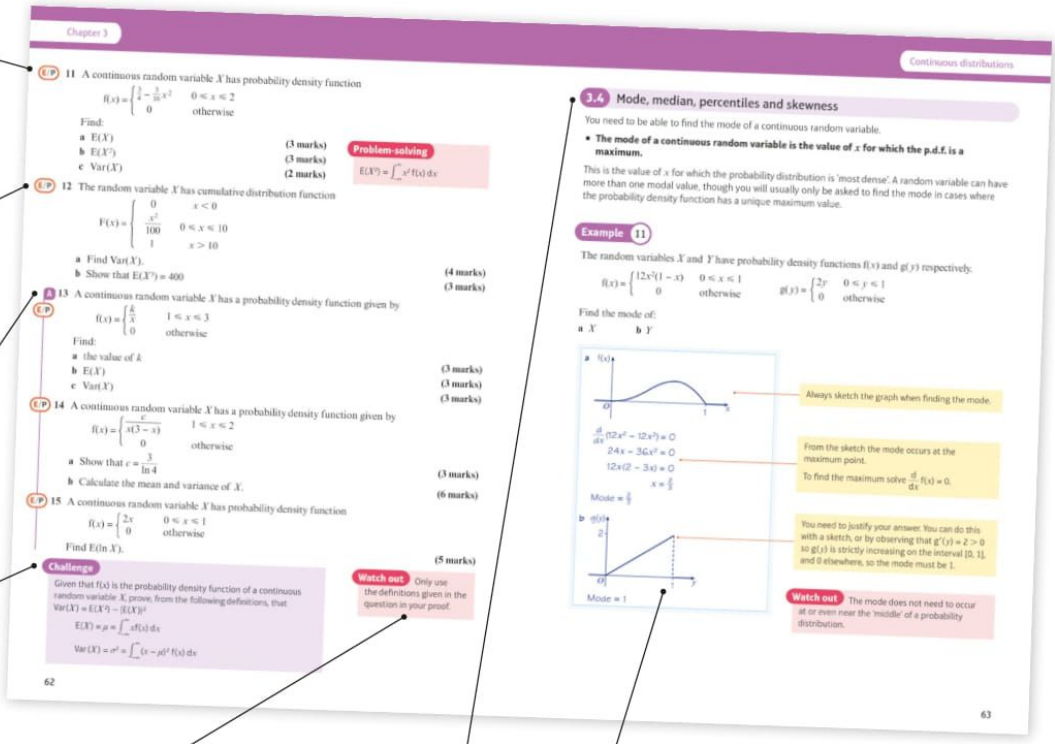
The *Prior knowledge check* helps make sure you are ready to start the chapter

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

A level content is clearly flagged

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



Chapter 3

Continuous distributions

11 A continuous random variable X has probability density function

$$f(x) = \begin{cases} \frac{3}{4} - \frac{3}{16}x^2 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $E(X)$ (3 marks)

b $E(X^2)$ (3 marks)

c $\text{Var}(X)$ (2 marks)

Problem-solving $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

12 The random variable X has cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{100} & 0 \leqslant x \leqslant 10 \\ 1 & x > 10 \end{cases}$$

a Find $\text{Var}(X)$. (4 marks)

b Show that $E(X^3) = 400$ (3 marks)

13 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{k}{x} & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the value of k (3 marks)

b $E(X)$ (3 marks)

c $\text{Var}(X)$ (3 marks)

14 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{c}{x(3-x)} & 1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $c = \frac{3}{\ln 4}$ (3 marks)

b Calculate the mean and variance of X. (6 marks)

15 A continuous random variable X has probability density function

$$f(x) = \begin{cases} 2x & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(\ln X)$. (5 marks)

Challenge

Given that $f(x)$ is the probability density function of a continuous random variable X, prove, from the following definitions, that $\text{Var}(X) = E(X^2) - [E(X)]^2$

$E(X) = \mu = \int_{-\infty}^{\infty} x f(x)\,dx$

$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

Watch out Only use the definitions given in the question in your proof.

62

3.4 Mode, median, percentiles and skewness

You need to be able to find the mode of a continuous random variable.

- **The mode of a continuous random variable is the value of x for which the p.d.f. is a maximum.**

This is the value of x for which the probability distribution is 'most dense'. A random variable can have more than one modal value, though you will usually only be asked to find the mode in cases where the probability density function has a unique maximum value.

Example 11

The random variables X and Y have probability density functions $f(x)$ and $g(y)$ respectively.

$$f(x) = \begin{cases} 12x^2(1-x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases} \qquad g(y) = \begin{cases} 2y & 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the mode of:

a X b Y

a

$\frac{d}{dx}(12x^2 - 12x^3) = 0$

$24x - 36x^2 = 0$

$12x(2 - 3x) = 0$

$x = \frac{2}{3}$

Mode $= \frac{2}{3}$

Always sketch the graph when finding the mode.

From the sketch the mode occurs at the maximum point.

To find the maximum solve $\frac{d}{dx} f(x) = 0$.

b

Mode = 1

You need to justify your answer. You can do this with a sketch, or by observing that $g'(y) = 2 > 0$ so $g(y)$ is strictly increasing on the interval $[0, 1]$, and 0 elsewhere, so the mode must be 1.

Watch out The mode does not need to occur at or even near the 'middle' of a probability distribution.

63

Exam-style questions are flagged with (E)

Problem-solving questions are flagged with (P)

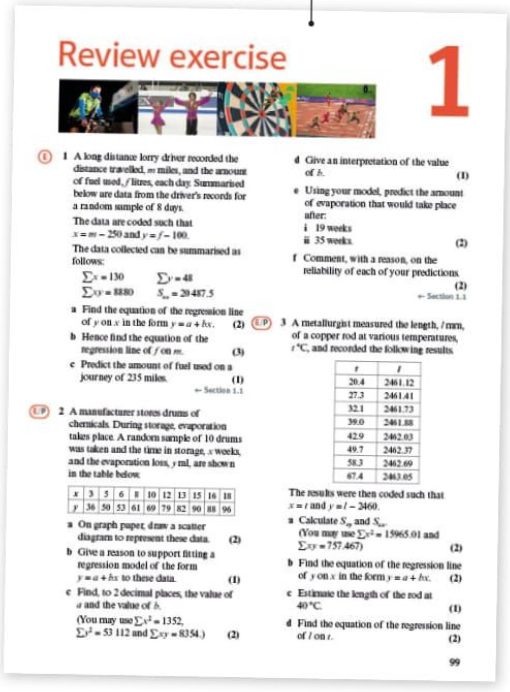
Problem-solving boxes provide hints, tips and strategies, and *Watch out* boxes highlight areas where students often lose marks in their exams

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

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

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



Review exercise 1

(E) **1** A long distance lorry driver recorded the distance travelled, m miles, and the amount of fuel used, f litres, each day. Summarised below are data from the driver's records for a random sample of 8 days.

The data are coded such that $x = m - 250$ and $y = f - 100$.

The data collected can be summarised as follows:

$\sum x = 130$ $\quad \sum y = 48$

$\sum xy = 8880$ $\quad S_{xx} = 29\,487.5$

a Find the equation of the regression line of y on x in the form $y = a + bx$. (2)

b Hence find the equation of the regression line of f on m. (3)

c Predict the amount of fuel used on a journey of 235 miles. (1)

← Section 1.1

(P) **2** A manufacturer stores drums of chemicals. During storage, evaporation takes place. A random sample of 10 drums was taken and the time in storage, x weeks, and the evaporation loss, y ml, are shown in the table below.

x	3	5	6	8	10	12	13	15	16	18
y	36	50	53	61	69	79	82	90	88	96

a On graph paper, draw a scatter diagram to represent these data. (2)

b Give a reason to support fitting a regression model of the form $y = a + bx$ to these data. (1)

c Find, to 2 decimal places, the value of a and the value of b.

(You may use $\sum x^2 = 1352$, $\sum y^2 = 53\,112$ and $\sum xy = 8354$.) (2)

d Give an interpretation of the value of b. (1)

e Using your model, predict the amount of evaporation that would take place after:

i 19 weeks

ii 35 weeks. (2)

f Comment, with a reason, on the reliability of each of your predictions. (2)

← Section 1.1

(P) **3** A metallurgist measured the length, l mm, of a copper rod at various temperatures, t °C, and recorded the following results.

t	l
20.4	2461.12
27.3	2461.41
32.1	2461.73
39.0	2461.88
42.9	2462.03
49.7	2462.37
58.3	2462.69
67.4	2463.05

The results were then coded such that $x = t$ and $y = l - 2460$.

a Calculate S_{xx} and S_{xy}.

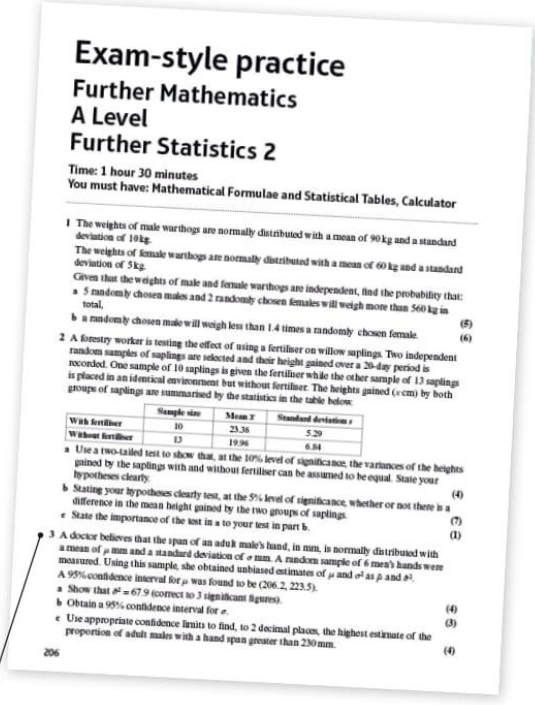
(You may use $\sum x^2 = 15965.01$ and $\sum xy = 757.467$) (2)

b Find the equation of the regression line of y on x in the form $y = a + bx$. (2)

c Estimate the length of the rod at 40 °C. (1)

d Find the equation of the regression line of l on t. (2)

99



Exam-style practice

Further Mathematics A Level Further Statistics 2

Time: 1 hour 30 minutes

You must have: Mathematical Formulae and Statistical Tables, Calculator

1 The weights of male warthogs are normally distributed with a mean of 90 kg and a standard deviation of 10 kg.

The weights of female warthogs are normally distributed with a mean of 60 kg and a standard deviation of 5 kg.

Given that the weights of male and female warthogs are independent, find the probability that:

a 5 randomly chosen males and 2 randomly chosen females will weigh more than 560 kg in total, (6)

b a randomly chosen male will weigh less than 1.4 times a randomly chosen female. (6)

2 A forestry worker is testing the effect of using a fertiliser on willow saplings. Two independent random samples of saplings are selected and their height gained over a 28-day period is recorded. One sample of 10 saplings is given the fertiliser while the other sample of 13 saplings is placed in an identical environment but without fertiliser. The heights gained (x cm) by both groups of saplings are summarised by the statistics in the table below.

	Sample size	Mean $\bar{x}$	Standard deviation s
With fertiliser	10	23.36	5.29
Without fertiliser	13	19.96	6.84

a Use a two-tailed test to show that, at the 10% level of significance, the variances of the heights gained by the saplings with and without fertiliser can be assumed to be equal. State your hypotheses clearly. (7)

b Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is a difference in the mean height gained by the two groups of saplings. (4)

c State the importance of the test in a to your test in part b. (1)

3 A doctor believes that the span of an adult male's hand, in mm, is normally distributed with a mean of μ mm and a standard deviation of σ mm. A random sample of 6 men's hands were measured. Using this sample, she obtained unbiased estimates of μ and σ^2 as $\hat{\mu}$ and $\hat{\sigma}^2$.

A 95% confidence interval for μ was found to be (206.2, 223.5).

a Show that $\hat{\sigma}^2 = 67.9$ (correct to 3 significant figures). (4)

b Obtain a 95% confidence interval for σ. (3)

c Use appropriate confidence limits to find, to 2 decimal places, the highest estimate of the proportion of adult males with a hand span greater than 230 mm. (4)

206

AS and A level practice papers at the back of the book help you prepare for the real thing.

Extra online content

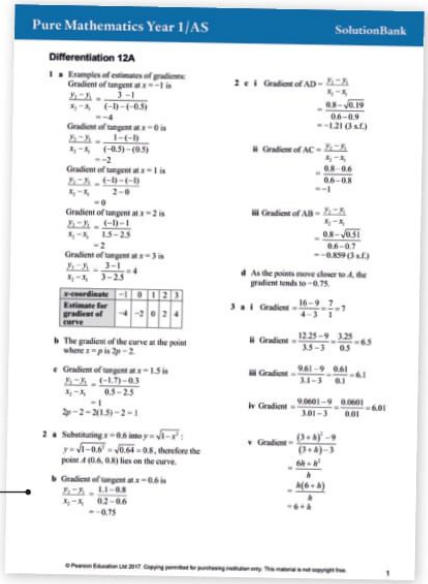
Whenever you see an *Online* box, it means that there is extra online content available to support you.

SolutionBank

SolutionBank provides a full worked solution for every question in the book.

Online Full worked solutions are available in SolutionBank.

Download all the solutions as a PDF or quickly find the solution you need online


Pure Mathematics Year 1/AS SolutionBank

Differentiation 12A

Use of technology

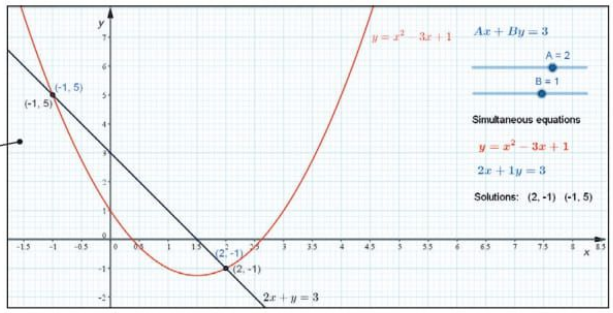
Explore topics in more detail, visualise problems and consolidate your understanding using pre-made GeoGebra activities.

Online Find the point of intersection graphically using technology.

GeoGebra

GeoGebra-powered interactives

Interact with the maths you are learning using GeoGebra's easy-to-use tools



Access all the extra online content for free at:

www.pearsonschools.co.uk/fs2maths

You can also access the extra online content by scanning this QR code:

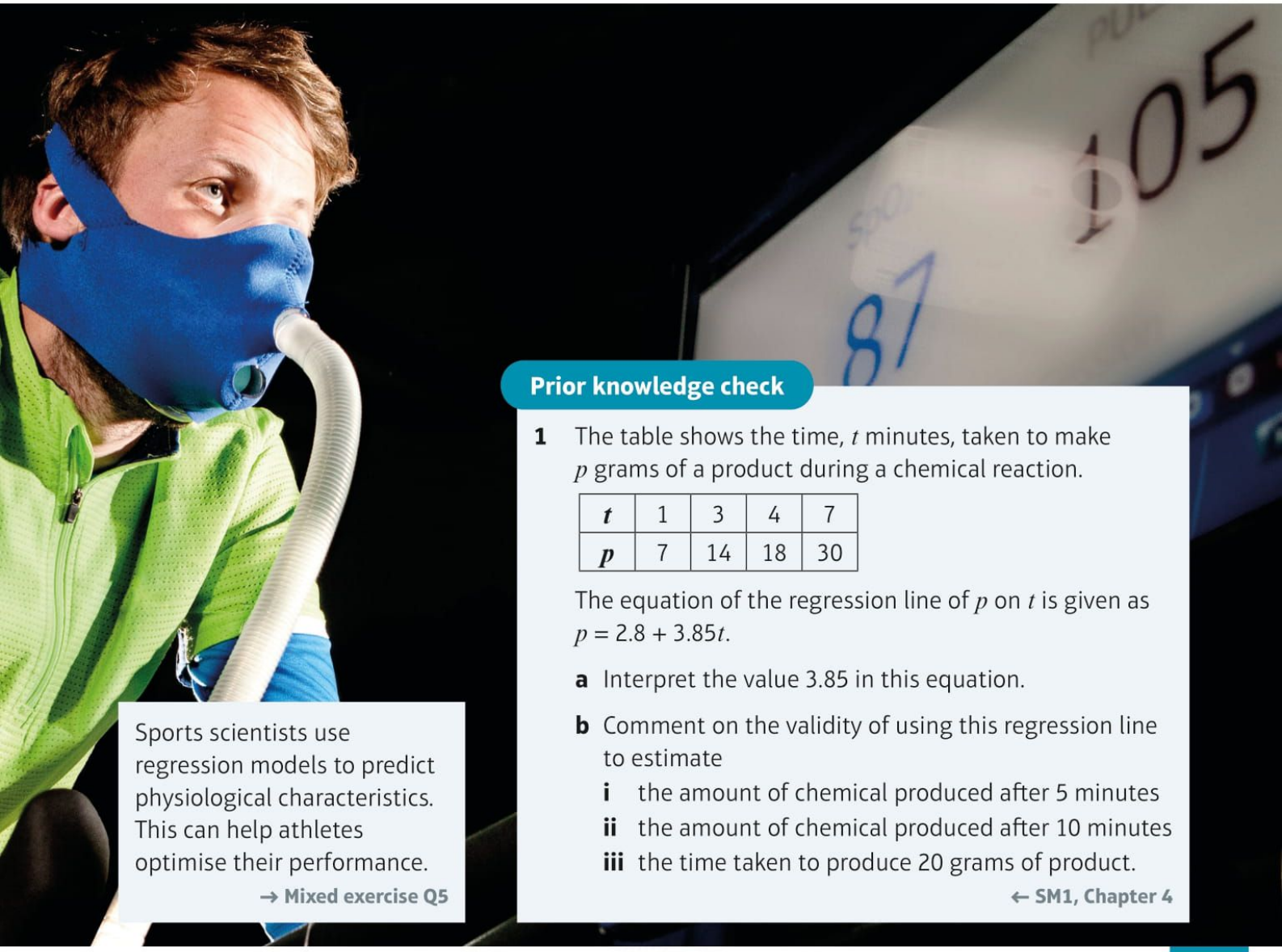
Linear regression

1

Objectives

After completing this chapter, you should be able to:

- Calculate the equation of a regression line using raw data or summary statistics → pages 2-8
- Use coding to find the equation of a regression line → pages 8-10
- Calculate residuals and use them to test for linear fit and identify outliers → pages 10-15
- Calculate the residual sum of squares (RSS) → pages 13-15



Sports scientists use regression models to predict physiological characteristics. This can help athletes optimise their performance.

→ Mixed exercise Q5

Prior knowledge check

1 The table shows the time, t minutes, taken to make p grams of a product during a chemical reaction.

t	1	3	4	7
p	7	14	18	30

The equation of the regression line of p on t is given as $p = 2.8 + 3.85t$.

a Interpret the value 3.85 in this equation.

b Comment on the validity of using this regression line to estimate

i the amount of chemical produced after 5 minutes

ii the amount of chemical produced after 10 minutes

iii the time taken to produce 20 grams of product.

← SM1, Chapter 4

1.1 Least squares linear regression

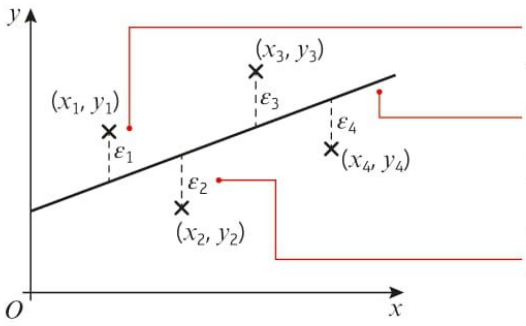
When you are analysing bivariate data, you can use a **least squares regression line** to predict values of the dependent (response) variable for given values of the independent (explanatory) variable. If the response variable is y and the explanatory variable is x, you should use the regression line of $\boldsymbol{y}$ **on** $\boldsymbol{x}$, which can be written in the form $y = a + bx$.

Links You should only use the regression line to make predictions for values of the dependent variable that are within the range of the given data. This is called **interpolation**. Making predictions for values outside of the range of the given data is called **extrapolation** and produces a less reliable prediction. ← SM1, Section 4.2

The least squares regression line is the line that minimises the **sum of the squares of the residuals** of each data point.

- **The residual of a given data point is the difference between the observed value of the dependent variable and the predicted value of the dependent variable.**

Notation The Greek letter epsilon (ε) is sometimes used to denote a residual.



ε_1 is the residual of the data point (x_1, y_1)

The least squares regression line of y on x is the straight line that minimises the value of $\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_2^4$. In general, if each data point has residual ε_i, the regression line minimises the value of $\sum \varepsilon_i^2$.

The observed value of the dependent variable, y_2, is **less** than the predicted value, so the residual of (x_2, y_2) will be **negative**.

You need to be able to find the equation of a **least squares regression line** using raw data or summary statistics.

- **The equation of the regression line of $\boldsymbol{y}$ on $\boldsymbol{x}$ is:**

$$y = a + bx$$

where $b = \frac{S_{xy}}{S_{xx}}$ **and** $a = \bar{y} - b\bar{x}$

Watch out You can calculate a and b directly from raw data using your calculator. However, you might be given summary statistics in the exam so you need to be familiar with these formulae.

S_{xy} and S_{xx} are known as **summary statistics** and you can calculate them using the following formulae:

- $S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

Example 1

The results from an experiment in which different masses were placed on a spring and the resulting length of the spring measured, are shown below.

Mass, x (kg)	20	40	60	80	100
Length, y (cm)	48	55.1	56.3	61.2	68

a Calculate S_{xx} and S_{xy}.
(You may use $\sum x = 300$ $\sum x^2 = 22000$ $\overline{x} = 60$ $\sum xy = 18238$ $\sum y^2 = 16879.14$ $\sum y = 288.6$ $\overline{y} = 57.72$)

b Calculate the regression line of y on x.

c Use your equation to predict the length of the spring when the applied mass is:
i 58 kg
ii 130 kg

d Comment on the reliability of your predictions.

Online Explore the calculation of a least squares regression line using GeoGebra.



a $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$
$= 22000 - \frac{300^2}{5}$
$= 4000$

$S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$
$= 18238 - \frac{300 \times 288.6}{5}$
$= 922$

Use the standard formulae to calculate S_{xx} and S_{xy}. Write down any formulae you are using before you substitute.

b $b = \frac{S_{xy}}{S_{xx}} = \frac{922}{4000} = 0.2305$
$a = \overline{y} - b\overline{x}$
$= 57.72 - 0.2305 \times 60$
$= 43.89$

Use the formulae to calculate a and b. If you want to check your answer using your calculator, make sure you use the correct mode for linear regression with bivariate data. On some calculators this mode is labelled $\boldsymbol{y = a + bx}$.

$y = 43.89 + 0.2305x$

Remember to write the equation at the end. The numbers should be given to a suitable degree of accuracy.

c i $y = 43.89 + 0.2305 \times 58$
$= 57.3$ cm (3 s.f.)
ii $y = 43.89 + 0.2305 \times 130$
$= 73.9$ cm (3 s.f.)

Substitute the given values into the equation of the regression line.

d Assuming the model is reasonable, the prediction when the mass is 58 kg is reliable since this is within the range of the data.

This is called **interpolation**.

The prediction when the mass is 130 kg is less reliable since this is outside the range of the data.

This is called **extrapolation**.

Example 2

A scientist working in agricultural research believes that there is a linear relationship between the amount of a certain food supplement given to hens and the hardness of the shells of the eggs they lay. As an experiment, controlled quantities of the supplement were added to the hens' normal diet for a period of two weeks and the hardness of the shells of the eggs laid at the end of this period was then measured on a scale from 1 to 10, with the following results:

Food supplement, f (g/day)	2	4	6	8	10	12	14
Hardness of shells, h	3.2	5.2	5.5	6.4	7.2	8.5	9.8

a Find the equation of the regression line of h on f.

(You may use $\sum f = 56$ $\sum h = 45.8$ $\bar{f} = 8$ $\bar{h} = 6.543$ $\sum f^2 = 560$ $\sum fh = 422.6$)

b Interpret what the values of a and b tell you.

a $S_{fh} = \sum fh - \dfrac{\sum f \sum h}{n}$

$= 422.6 - \dfrac{56 \times 45.8}{7} = 56.2$

$S_{ff} = \sum f^2 - \dfrac{(\sum f)^2}{n}$

$= 560 - \dfrac{56^2}{7} = 112$

$b = \dfrac{S_{fh}}{S_{ff}} = \dfrac{56.2}{112}$

$= 0.5017\ldots$ hardness units per g per day

$a = \bar{h} - b\bar{f}$

$= 6.543 - 0.5017\ldots \times 8$

$= 2.5287\ldots$ hardness units

$h = 2.53 + 0.502f$

b a estimates the shell strength when no supplement is given (i.e. when $f = 0$). Zero is only just outside the range of f so it is reasonable to use this value.
b estimates the rate at which the hardness increases with increased food supplement; in this case for every extra one gram of food supplement per day the hardness increases by 0.502 (3 s.f.) hardness units.

Watch out The variables given might not be x and y. Be careful that you use the correct values when you substitute into the formulae. It can sometimes help to write x next to the explanatory variable in the table (f) and y next to the response variable (h).

When dealing with a real problem do not forget to put the units of measurement for the two constants.

Make sure you give your answer in the context of the question. Don't just say that one value increases as the other increases – you need to comment on the **rate** of increase of hardness.

Example 3

A repair workshop finds it is having a problem with a pressure gauge it uses. It decides to have it checked by a specialist firm. The following data were obtained.

Gauge reading, x (bars)	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8
Correct reading, y (bars)	0.96	1.33	1.75	2.14	2.58	2.97	3.38	3.75

(You may use $\sum x = 19.2$ $\sum x^2 = 52.8$ $\sum y = 18.86$ $\sum y^2 = 51.30$ $\sum xy = 52.04$)

a Show that $S_{xy} = 6.776$ and find S_{xx}.

It is thought that a linear relationship of the form $y = a + bx$ could be used to describe these data.

b Use linear regression to find the values of a and b giving your answers to 3 significant figures.

c Draw a scatter diagram to represent these data and draw the regression line on your diagram.

d The gauge shows a reading of 2 bars. Using the regression equation, work out what the correct reading should be.

a $S_{xy} = \sum xy - \dfrac{\sum x \sum y}{n}$

$= 52.04 - \dfrac{19.2 \times 18.86}{8} = 6.776$

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$

$= 52.8 - \dfrac{(19.2)^2}{8} = 6.72$

Quote the formula you are going to use before substituting the values.

b $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{6.776}{6.72} = 1.0083\ldots$

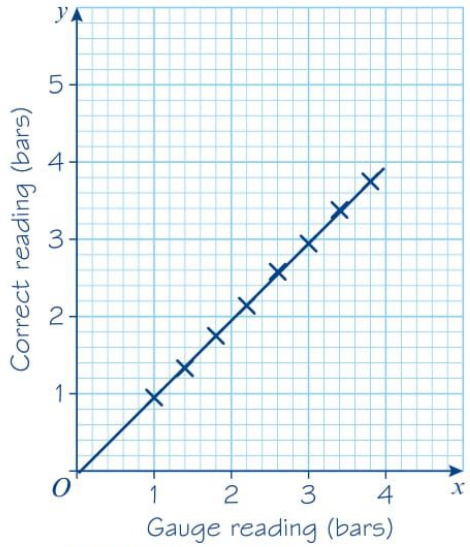
$a = \bar{y} - b\bar{x} = \dfrac{18.86}{8} - 1.0083\ldots \times \dfrac{19.2}{8}$

$= -0.0625$

Regression line is: $y = -0.0625 + 1.008x$

or $y = 1.008x - 0.0625$

c



To draw the regression line either plot the point $(0, a)$ and use the gradient or find two points on the line.

In this case using $x = 1$ gives $y = 0.95$ and using $x = 3$ gives $y = 2.96$.

d $y = (1.008 \times 2) - 0.0625$

$y = 1.95$ bar (3 s.f.)

Exercise 1A

1 The equation of a regression line in the form $y = a + bx$ is to be found. Given that $S_{xx} = 15$, $S_{xy} = 90$, $\overline{x} = 3$ and $\overline{y} = 15$, work out the values of a and b.

2 Given that $S_{xx} = 30$, $S_{xy} = 165$, $\overline{x} = 4$ and $\overline{y} = 8$, find the equation of the regression line of y on x.

3 The equation of a regression line is to be found. The following summary data is given:

$S_{xx} = 40$ $S_{xy} = 80$ $\overline{x} = 6$ $\overline{y} = 12$

Find the equation of the regression line in the form $y = a + bx$.

4 Data is collected and summarised as follows:

$\sum x = 10$ $\sum x^2 = 30$ $\sum y = 48$ $\sum xy = 140$ $n = 4$

a Work out $\overline{x}$, $\overline{y}$, S_{xx} and S_{xy}.

b Find the equation of the regression line of y on x in the form $y = a + bx$.

5 For the data in the table,

x	2	4	5	8	10
y	3	7	8	13	17

Hint Check your answer using the statistical functions on your calculator.

a calculate S_{xx} and S_{xy}

b find the equation of the regression line of y on x in the form $y = a + bx$.

(P) **6** Research was done to see if there is a relationship between finger dexterity and the ability to do work on a production line. The data is shown in the table.

Dexterity score, x	2.5	3	3.5	4	5	5	5.5	6.5	7	8
Productivity, y	80	130	100	220	190	210	270	290	350	400

The equation of the regression line for these data is $y = -59 + 57x$.

a Use the equation to estimate the productivity of someone with a dexterity of 6.

b Give an interpretation of the value of 57 in the equation of the regression line.

c State, giving in each case a reason, whether or not it would be reasonable to use this equation to work out the productivity of someone with dexterity of:

i 2 **ii** 14

7 A field was divided into 12 plots of equal area. Each plot was fertilised with a different amount of fertilizer (h). The yield of grain (g) was measured for each plot. Find the equation of the regression line of g on h in the form $g = a + bh$ given the following summary data.

$\sum h = 22.09$ $\sum g = 49.7$ $\sum h^2 = 45.04$ $\sum g^2 = 244.83$ $\sum hg = 97.778$ $n = 12$

E/P **8** A mail order company pays for postage of its goods partly by destination and partly by total weight sent out on a particular day. The number of items sent out and the total weights were recorded over a seven-day period. The data are shown in the table.

Number of items, n	10	13	22	15	24	16	19
Weight, w (kg)	2800	3600	6000	3600	5200	4400	5200

a Use the coding $x = n - 10$ and $y = \frac{w}{400}$ to work out S_{xy} and S_{xx}. **(4 marks)**

b Work out the equation of the regression line for y on x. **(3 marks)**

c Work out the equation of the regression line for w on n. **(2 marks)**

d Use your regression equation to estimate the weight of 20 items. **(2 marks)**

e State why it would be unwise to use the regression equation to estimate the weight of 100 items. **(1 mark)**

f Use your equation of the regression line found in part **b** to work out the residuals of the coded data points (x, y). **(2 marks)**

g Use your equation of the regression line found in part **c** to work out the residuals of the original data points (n, w). **(2 marks)**

h Explain how your answers to parts **f** and **g** are related to the coding used. **(1 mark)**

E/P **9** The table shows the time, t hours, against the temperature, T°C, of a chemical reaction.

t	2	3	5	6	7	9	10
T	72	68	59	54	50	42	38

Given that the equation of the regression line of T on t is $T = 80.445 - 4.289t$,

a calculate the residual values. **(2 marks)**

b State, with a reason, whether a linear model is suitable in this case. **(1 mark)**

Given that $S_{tt} = 52$, $S_{TT} = 957.43$ and $S_{tT} = -223$,

c calculate the residual sum of squares (RSS). **(2 marks)**

A second chemical reaction has a RSS of 0.8754.

d State, with a reason, which reaction is most likely to have a linear fit. **(1 mark)**

E/P **10** A meteorologist is developing a model to describe the relationship between the number of hours of sunshine, s, and the daily rainfall, $f$mm, in summer.

A random sample of the number of hours sunshine and the daily rainfall is taken from 8 days and are summarised below:

$\sum s = 53.4$ $\sum s^2 = 395.76$ $\sum f = 29.9$ $\sum f^2 = 131.93$ $\sum sf = 171.66$

a Calculate S_{ss} and S_{sf}. **(2 marks)**

b Find the equation of the regression line of f on s. **(3 marks)**

c Use your equation to estimate the daily rainfall when there is 7.5 hours of sunshine. **(1 mark)**

d Calculate the residual sum of squares (RSS). **(3 marks)**

The table shows the residual for each value of s.

s	3.1	4.2	5.4	6.2	7.1	8.8	9.1	9.5
Residual	−0.177	−0.196	0.256	0.124	x	−0.129	−0.216	−0.032

e Find the value of x. **(2 marks)**

f By considering the signs of the residuals, explain whether or not the linear regression model is suitable for these data. **(1 mark)**

E/P **11** A random sample of 9 baby southern hairy-nosed wombats was taken. The age, x, in days, and the mass, y grams, was recorded. The results were as follows:

x	2	3	4	5	6	7	8	9	10
y	4	5	7	8	9	11	12	11	15

(You may use $S_{xx} = 60$ $S_{yy} = 98.89$ $S_{xy} = 75$)

a Find the equation of the regression line of y on x in the form $y = a + bx$ as a model for these results. Give the values of a and b correct to three significant figures. **(2 marks)**

b Show that the residual sum of squares is 5.14 to three significant figures. **(2 marks)**

c Calculate the residual values. **(2 marks)**

d Write down the outlier. **(1 mark)**

e **i** Comment on the validity of ignoring this outlier.

ii Ignoring the outlier, produce another model.

iii Use this model to estimate the mass of a baby wombat after 20 days.

iv Comment, giving a reason, on the reliability of your estimate. **(5 marks)**

E/P **12** The annual turnover, £t million of eight randomly selected UK companies, and the number of staff employed in 100s, s, is recorded and the data shown in the table below:

t, £million	1.2	1.5	1.8	2.1	2.5	2.7	2.8	3.1
s, 00s	1.1	1.4	1.7	2.2	2.4	2.6	2.9	3.2

($\sum t = 17.7$ $\sum s = 17.5$ $\sum t^2 = 42.33$ $\sum s^2 = 42.07$ $\sum ts = 42.16$)

a Calculate the equation of the regression line of s on t, giving your answer in the form $s = a + bt$. Give the values of a and b correct to three significant figures. **(3 marks)**

b Use your regression line to predict the number of employees in a UK company with an annual turnover of £2 300 000. **(2 marks)**

The table shows the residuals for each value of t:

t	1.2	1.5	1.8	2.1	2.5	2.7	2.8	3.1
Residual	0.0121	−0.0137	−0.0395	0.1347	−0.0997	p	0.0745	0.0487

c Find the value of p. **(2 marks)**

d By considering the signs of the residuals, or otherwise, comment on the suitability of the linear regression model for these data. **(1 mark)**

e Calculate the residual sum of squares (RSS). **(2 marks)**

A random sample of equivalent companies in France is taken and the residual sum of squares is found to be 0.421.

f State, with a reason, which sample is likely to have the better linear fit. **(1 mark)**

Pearson Edexcel AS and A level Further Mathematics

Further Statistics 2

Series Editor: Harry Smith

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

This book covers all the content needed for the optional **Pearson Edexcel AS and A level Further Statistics 2 exams**.

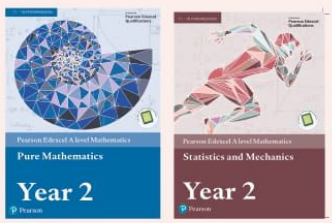
- Fully updated to match the 2017 specifications, with more of a focus on problem-solving and modelling.
- FREE additional online content to support your independent learning, including full worked solutions for every question in the book (SolutionBank), GeoGebra interactives.
- Includes access to an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

Pearson Edexcel AS and A level Mathematics books

Year 1/AS

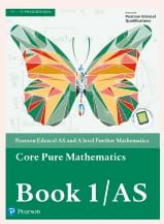


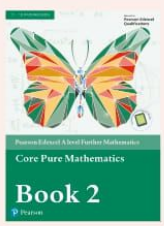
Year 2



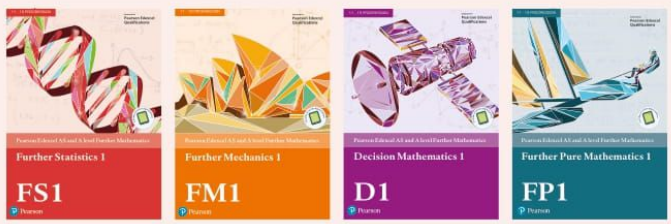
Pearson Edexcel AS and A level Further Mathematics books

Compulsory





Options





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

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

ISBN 978-1-292-18338-1

Pearson