



eClassroom

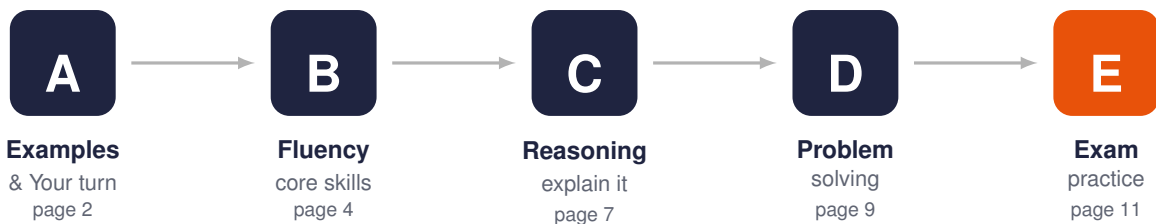
GCSE MATHEMATICS • DATA & STATISTICS

Scatter Graphs

Worksheet

Pearson Edexcel GCSE (1MA1) & International GCSE (4MA1) • Calculator allowed

Free Scatter Graphs worksheet for GCSE and iGCSE Maths, with full worked solutions



Tap a section to jump straight to it

By the end of this worksheet you can: plot points and describe correlation • spot outliers • draw a line of best fit and use it to estimate • know when an estimate is reliable, and that correlation does not prove causation • *Higher:* the equation of a line of best fit, its gradient in context, and the mean point.

Which questions are for me?

FOUNDATION Everyone starts here. These skills appear on both tiers.

HIGHER The same skills where Higher papers use them: the equation of a line of best fit, the mean point, residuals and comparing trends. Grades 6 to 9.

Every section has both, clearly labelled. Questions get harder as you go, and every skill in Sections A to D is needed in Section E.

Track your progress

Section	B	C	D	E
Foundation marks	23	10	13	15
Higher marks	16	14	17	15
My score				

Full worked solutions

Data & Statistics revision engine

Free courses

Formula sheet builder

Name

Class

Date



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

Page 1 of 14

A

Worked examples & Your turn

Read each example carefully. Cover it up, then answer the question beside it using the same method.

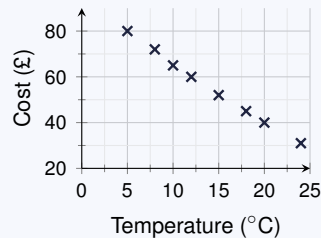
FOUNDATION

Everyone starts here

Worked example A1

Describe the correlation and what it shows.

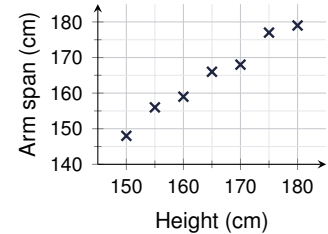
The points go down from left to right.



negative: warmer days, lower heating cost

Your turn A1

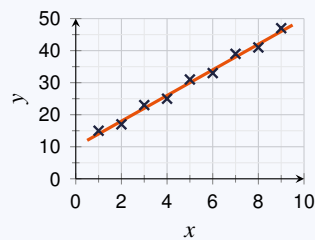
Describe the correlation and what it shows.



Worked example A2

Use the line of best fit to estimate y when $x = 6$

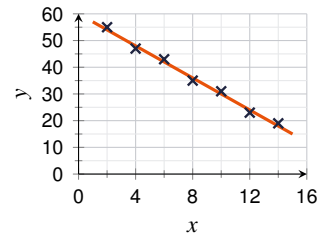
Go up from $x = 6$ to the line, then across to the y -axis.



34

Your turn A2

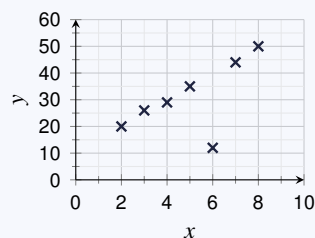
Use the line of best fit to estimate x when $y = 30$



Worked example A3

Which point is an outlier?

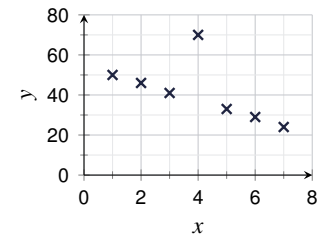
It does not follow the trend of the other points.



(6, 12)

Your turn A3

Which point is an outlier?



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

HIGHER*Higher strategies: equations of lines of best fit and the mean point***Worked example A4: a best-fit equation**

A line of best fit passes through (2, 30) and (8, 60). Find its equation.

$$\text{Gradient} = \frac{60 - 30}{8 - 2} = 5. \text{ Then } 30 = 5 \times 2 + c$$

$$y = 5x + 20$$

Your turn A4

A line of best fit passes through (0, 90) and (10, 40). Find its equation.

Worked example A5: the gradient in context

A line of best fit for plant height h cm after w weeks is $h = 1.5w + 4$. What do the 1.5 and the 4 mean?

The gradient is the change in h for each extra week.

grows 1.5 cm a week; about 4 cm at the start

Your turn A5

Ice-lolly sales S on a day with temperature t °C are modelled by $S = 12t - 50$. What does the 12 mean? Does the -50 make sense?

Worked example A6: the mean point

Find the mean point of (2, 10), (4, 14), (6, 19), (8, 21), (10, 26)

Mean of the x -values, and mean of the y -values: $30 \div 5$ and $90 \div 5$

(6, 18)

Your turn A6

Find the mean point of (1, 40), (3, 34), (5, 31), (7, 25), (9, 20)

Worked example A7: reliable estimates

The data runs from $x = 2$ to $x = 10$. Which estimate is more reliable: at $x = 6$ or at $x = 20$?

$x = 6$ is inside the data (interpolation). $x = 20$ is outside it (extrapolation): the trend may not continue.

$x = 6$

Your turn A7

A line of best fit uses children aged 4 to 12. Is it sensible to use it to estimate the height of a 30-year-old? Explain.



B**Fluency**

Each question is a designed sequence. From one part to the next, ask yourself: what changed, and what did it do to the answer?

FOUNDATION*Everyone starts here*

B1 What type of correlation would you expect between each pair? Choose positive, negative or none.

(a) height and arm span

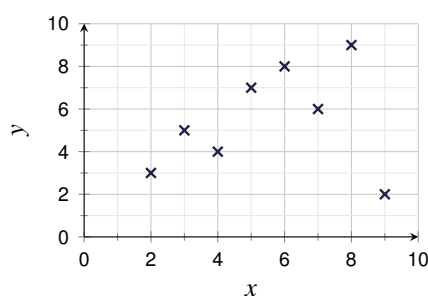
(b) the age of a car and its value

(c) shoe size and test mark

(d) temperature and hot-chocolate sales

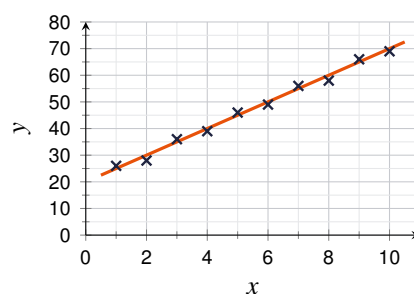
[4 marks]

- B2** (a) How many points are plotted?
(b) Write down the coordinates of the highest point.
(c) Write down the coordinates of the outlier.
(d) Describe the correlation.



[4 marks]

- B3** Use the line of best fit.
- (a) Estimate y when $x = 4$
(b) Estimate y when $x = 7$
(c) Estimate x when $y = 50$
(d) Would an estimate at $x = 15$ be reliable? Explain.



[4 marks]



HIGHER*continued: gradients, mean points and predictions*

B8 Each equation is a line of best fit. Explain what the gradient and the intercept mean.

(a) Journey time $T = 2d + 20$ minutes, for a distance d km

(b) Test mark $M = -4w + 80$, where w is hours of gaming the night before

(c) Height $H = 6a + 50$ cm for a child aged a years

(d) Car value $V = 20\,000 - 1500t$ pounds after t years

[4 marks]

B9 Find the mean point of each set of data.

(a) (1, 5), (3, 9), (5, 13) (b) (2, 20), (4, 16), (6, 15), (8, 9)

(c) (10, 3), (20, 7), (30, 8), (40, 12), (50, 15) (d) (1.5, 40), (2.5, 35), (3.5, 24), (4.5, 21)

[4 marks]

B10 A line of best fit is $y = 3x + 8$, from data with x between 1 and 12

(a) Predict y when $x = 5$

(b) Predict y when $x = 9$

(c) Find x when $y = 29$

(d) Comment on predicting y when $x = 40$

[4 marks]



C**Reasoning**

Show your working and explain your thinking. A correct answer without a reason does not score full marks.

FOUNDATION

Everyone starts here

- C1** Tom says, "Positive correlation means all the points lie exactly on a straight line going up."
Explain why Tom is wrong. [2 marks]
- C2** Sam draws his line of best fit by joining the first point to the last point.
Explain why this may not give a good line of best fit. [2 marks]
- C3** A scatter graph shows strong positive correlation between ice-cream sales and the number of shark sightings at a beach.
Does buying ice cream attract sharks? Explain, and suggest a better reason for the correlation. [3 marks]
- C4** Kim draws her line of best fit so that it passes through the origin, although none of her points is near there.
Explain why this may be a mistake. [3 marks]



HIGHER*Higher tier only*

- C5** Jo's data runs from $x = 0$ to $x = 10$. She uses her line of best fit to predict the value at $x = 30$
Explain why her prediction may be unreliable. [3 marks]
- C6** Explain why a good line of best fit should pass through the mean point of the data. [3 marks]
- C7** Two students draw different lines of best fit for the same data. Describe how you could decide which line fits the data better. [3 marks]
- C8** The line $y = mx + c$ passes through the mean point $(\bar{x}, \bar{y})$ of n points $(x_1, y_1), \dots, (x_n, y_n)$
- (a) Show that $\bar{y} = m\bar{x} + c$
- (b) Hence show that the vertical distances from the points to the line, $y_i - (mx_i + c)$, add up to 0 [5 marks]



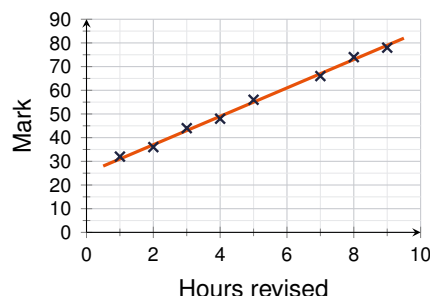
D**Problem solving**

You decide the method. Read the whole question first, and show every step.

FOUNDATION

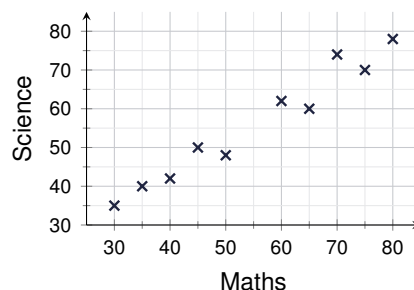
Everyone starts here

- D1** The graph shows hours revised and test marks.
The line of best fit is drawn.
Ali revised for 6 hours. Estimate his mark, and say whether your estimate is reliable.



[3 marks]

- D2** The graph shows the maths and science marks of 10 students.
(a) Describe the relationship.
(b) A new student scored 55 in maths. Draw a line of best fit and estimate their science mark.
(c) The teacher says, "Every student who is good at maths is good at science." Comment.



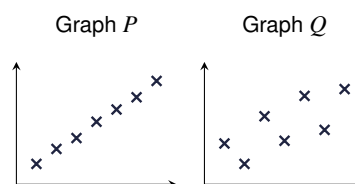
[4 marks]

- D3** A scatter graph shows the age and value of some used cars. One car is shown far above the line of best fit.

Give two possible reasons why this car is worth more than the trend suggests.

[3 marks]

- D4** Which graph shows the stronger correlation?
Explain.



[3 marks]

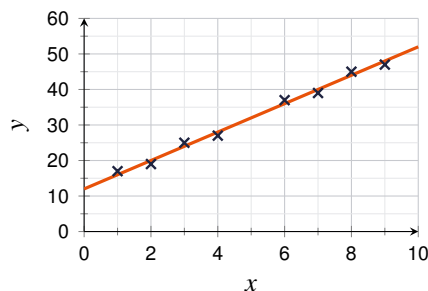


HIGHER*Higher tier only*

D5 The line of best fit is drawn.

(a) Find its equation.

(b) Use your equation to estimate y when $x = 7$



[4 marks]

D6 A line of best fit is $y = 4x + 15$. A point is flagged as an outlier if its y -value is more than 8 away from the line.

(a) At $x = 5$ the actual value is 30. Is this point an outlier?

(b) At $x = 8$ the actual value is 60. Is this point an outlier?

[4 marks]

D7 Two groups of plants are grown with different fertilisers. The lines of best fit for height h cm after w weeks are $h = 2w + 10$ for A and $h = 3w + 4$ for B

After how many weeks are the two groups expected to be the same height? Which group is taller after that?

[4 marks]

D8 Five points are $(2, 9)$, $(4, 13)$, $(6, k)$, $(8, 21)$ and $(10, 27)$. Their line of best fit is $y = 2x + 5$, and it passes through the mean point.

(a) Find $\bar{x}$, and explain why $\bar{y} = 17$

(b) Find k .

(c) How far is the point $(6, k)$ from the line, measured vertically?

[5 marks]



E**Exam practice**

Exam-style questions. Write your answers in the spaces. Aim for about one minute per mark.

FOUNDATION

Foundation and Higher crossover

- E1** The scatter graph shows the hours some students revised and their test marks.

(a) Describe the correlation.

.....
(1)

(b) Write down the coordinates of the outlier.

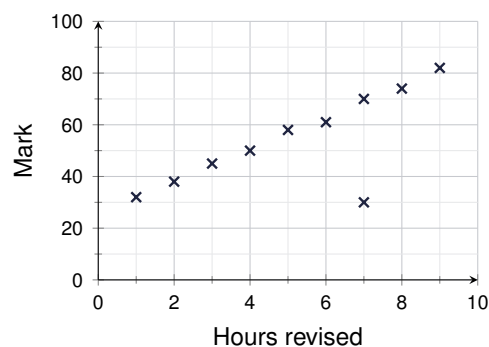
.....
(1)

(c) Draw a line of best fit.

(1)

(d) Estimate the mark for 5.5 hours of revision.

.....
(1)



(Total for Question E1 is 4 marks)

- E2** Using the graph in E1, Kim estimates the mark for a student who revised for 20 hours. Explain why her estimate may not be reliable.

(Total for Question E2 is 2 marks)

- E3** Five points are plotted. Here are three more.

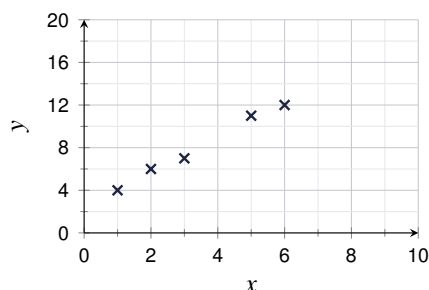
x	4	7	8
y	9	15	16

(a) Plot the three points.

(2)

(b) Describe the relationship between x and y .

.....
(1)



(Total for Question E3 is 3 marks)



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

E4 A newspaper says, "Towns with more coffee shops have more car crashes, so coffee shops cause crashes."

(a) What type of correlation does the newspaper describe?

.....

(1)

(b) Explain why the newspaper's conclusion may be wrong.

(2)

(Total for Question E4 is 3 marks)

E5 The graph shows the age and value of some cars.

(a) Draw a line of best fit.

(1)

(b) Estimate the value of a car that is 4 years old.

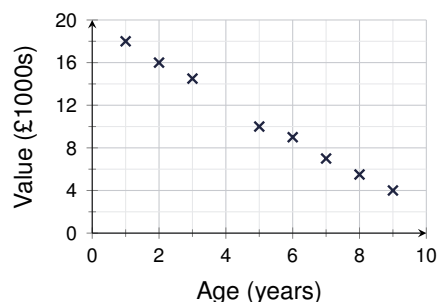
.....

(1)

(c) Would your line give a reliable value for a 15-year-old car? Explain.

.....

(1)



(Total for Question E5 is 3 marks)

HIGHER

Higher tier only

E6 A line of best fit for the marks M of students who revised for h hours passes through (0, 20) and (10, 70).

(a) Find the equation of the line.

.....

(2)

(b) Explain what the gradient tells you.

(1)

(Total for Question E6 is 3 marks)



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

- E7** The mean point of some data is (6, 38). The line of best fit passes through the mean point and has a gradient of 4

(a) Find the equation of the line.

.....
(2)

(b) Estimate y when $x = 9$

.....
(1)

(Total for Question E7 is 3 marks)

- E8** The table shows 8 pairs of values.

x	2	3	4	5	6	7	8	9
y	11	13	15	17	20	22	28	10

(a) Which pair is an outlier?

.....
(1)

(b) Leaving out the outlier, find the mean point.

.....
(2)

(c) Which estimate would be more reliable: y when $x = 5.5$ or y when $x = 20$? Explain.

(1)

(Total for Question E8 is 4 marks)



E9 A line of best fit passes through (2, 14) and (10, 54)

(a) Find the equation of the line.

.....
(2)

(b) Estimate y when $x = 7$

.....
(1)

(c) Sam says, "If x doubles, y doubles." Use your equation to show that Sam is wrong.

(2)

(Total for Question E9 is 5 marks)

Finished? Mark your work with the **full worked solutions**, fill in your scores on the front page, then keep the skill sharp with unlimited questions on the **Data & Statistics revision engine**.



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk