



Lines of best fit and regression

About 50 minutes Calculator allowed 44 marks

Date _____

A Do Now

The first question is what today's lesson needs. The others bring back earlier topics.

- A1** Needed today [2 marks] **A2** From chapter 5 [2 marks]
- A scatter diagram shows trunk width against age of a tree for 19 trees. The points go up from left to right but are widely spread out. Describe the correlation and interpret it in the context of the question.
- For the lengths, in mm, of some worms, none are 20 or less and the cumulative frequency is 11 at 30, 37 at 40, 65 at 50, 81 at 60. Estimate how many worms had 55 mm or less.

Answer

- A3** From chapter 8 [2 marks] **A4** Maths skill [2 marks]
- In Dunmere the death rates per 1000 are 2.4 for people under 45, 6 for people aged 45–64 and 58 for people aged 65 and over. A standard population is 40% under 45, 35% aged 45–64 and 25% aged 65 and over. Work out the standardised death rate for Dunmere.
- Write $\frac{28}{168}$ in its simplest form.

Answer

Answer

B Examples, then your turn

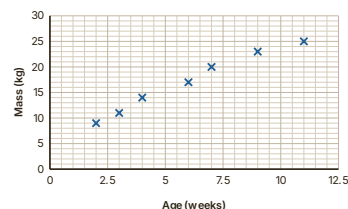
Your teacher works through each example with you. Then try the one beside it.

EXAMPLE 1

A shop records the price, x pounds, and the number sold, y , of six items. x : 2, 4, 5, 7, 8, 10. y : 31, 38, 44, 50, 57, 68. Work out the double mean point. How is it used to draw the line of best fit?

Hint The double mean point is $(\bar{x}, \bar{y})$: the mean of the x values and the mean of the y values.

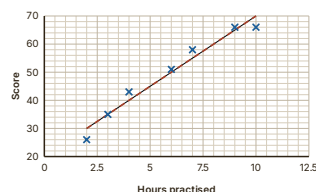
- B1** The scatter diagram shows the ages, 2, 3, 4, 6, 7, 9, 11 weeks, and the masses, 9, 11, 14, 17, 20, 23, 25 kg, of 7 lambs. Work out the double mean point. Plot it and draw the line of best fit. [3 marks]



Answer

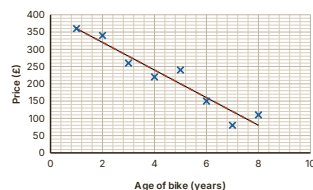
EXAMPLE 2

The line of best fit is drawn on the scatter diagram. Find its equation in the form $y = a + bx$.



Hint Read two points on the line that are far apart. Use the numbers on the scales.

- B2** The line of best fit is drawn on the scatter diagram. Find its equation in the form $y = a + bx$. [3 marks]



Answer

EXAMPLE 3

A taxi firm's fares are modelled by $y = 2.5 + 1.8x$, where x is the length of a journey in miles and y is the fare in pounds. Interpret 2.5 and 1.8 in context.

Hint a is the value of y when $x = 0$. b is how much y goes up for each 1 more of x .

- B3** The salaries of teachers at a school are modelled by $y = 28 + 1.2x$, where x is the number of years they have taught and y is the salary in thousands of pounds. Interpret 1.2 and 28 in context. [2 marks]

.....
.....

EXAMPLE 4

For 10 cars aged 1 to 8 years, the line of best fit is $y = 15 - 1.5x$, where x is the age in years and y is the value in thousands of pounds. Estimate the value of a car aged 5 years and of a car aged 12 years. Comment on how reliable each estimate is.

Hint Is the age inside the range of the data, 1 to 8 years, or outside it?

B4 A farmer tests amounts of fertiliser from 10 g to 50 g on tomato plants. The line of best fit is $y = 1.6 + 0.05x$, where x is the fertiliser in grams and y is the crop in kilograms. [3 marks]

Estimate the crop for 30 g and for 200 g. Comment on the reliability of each estimate.

EXAMPLE 5

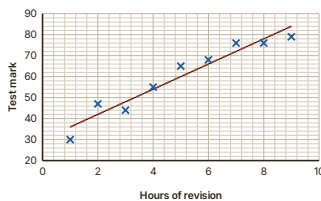
For some apples, the line of best fit between length, x cm, and mass, y g, is $y = a + 9x$. The double mean point is (8, 110). Work out a . Does a have a meaning in context?

Hint The line passes through $(\bar{x}, \bar{y})$, so put $x = 8$ and $y = 110$ into the equation.

B5 For some dogs, the line of best fit between mass, x kg, and the food they eat each day, y g, is $y = a + 15x$. The double mean point is (20, 380). Work out a . Does a have a meaning in context? Give a reason. [3 marks]

c Practice

- C1** The line of best fit is drawn on the scatter diagram. Use it to estimate the test mark of a student who revised for 6.5 hours. [1 mark]



Answer

- C2** The line of best fit for some data has equation $y = a + 2.4x$. The double mean point is (15, 50). Work out the value of a . [2 marks]

Answer

- C3** For a type of fish, the line of best fit between length, x cm, and mass, y g, is $y = -150 + 25x$. The fish in the data are 12 cm to 30 cm long. Give a reason why -150 has no meaning in context. Interpret 25. [2 marks]

- C4** Some data show positive correlation and have double mean point $(24, 61)$. Which of $y = 13 + 2x$, $y = 85 - x$ and $y = 61 + 2x$ could be the line of best fit? Give reasons. [2 marks]

D Exam-style questions

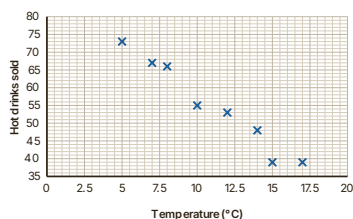
- D1** Leo measured the heights of children aged 2 to 10. His line of best fit is $y = 80 + 6x$, where x is the age in years and y is the height in cm. He says, "So a 40-year-old will be 320 cm tall." [2 marks]

Is Leo's estimate reliable? Give a reason for your answer.

- D2** Two classes practise typing. For class P the line of best fit is $y = 18 + 2.5x$ and for class Q it is $y = 25 + 1.5x$, where x is the number of weeks of practice and y is the typing speed in words per minute. Which class's speed increases faster? After how many weeks do the two lines give the same speed, and what is that speed? [3 marks]

Answer

- D3** A café records the midday temperature and the number of hot drinks sold on 8 days. The temperatures are 5, 7, 8, 10, 12, 14, 15, 17 °C and the drinks sold are 73, 67, 66, 55, 53, 48, 39, 39. The equation of the line of best fit is $y = 88 - 3x$.



- (a) Work out the double mean point. Show that it lies on the line $y = 88 - 3x$. [2 marks]

Answer

- (b) Interpret the gradient of the line in context. [1 mark]

- (c) Use the equation to estimate the number of hot drinks sold on a day when the temperature is 9 °C. [2 marks]

Answer

- (d) Explain why the equation should not be used to estimate the drinks sold when it is 30 °C. [2 marks]

Total for D3: 7 marks

E Extension

No route is given. The hints are at the end of the sheet. Use one at a time.

- E1** Five pairs of data are (3, 20), (5, 26), (8, k), (11, 41) and (13, 47). The line of best fit $y = 11 + 2.8x$ passes through the double mean point. Work out the value of k . [3 marks]

Answer

Extension hints

Cover these with paper. Uncover one at a time.

- E1** **Hint 1** The line of best fit passes through $(\bar{x}, \bar{y})$. Work out $\bar{x}$ first.
- Hint 2** $\bar{x} = 40 \div 5 = 8$, so $\bar{y} = 11 + 2.8 \times 8$.
- Hint 3** $\bar{y} = 33.4$, so the five y values add up to $5 \times 33.4 = 167$.

How did it go?

Colour one circle on each line: not yet, nearly, yes.

- I can draw a line of best fit through the double mean point. ☐ ☐ ☐
- I can find and interpret a and b in $y = a + bx$ in context. ☐ ☐ ☐
- I can say whether an estimate is reliable: interpolation or extrapolation. ☐ ☐ ☐