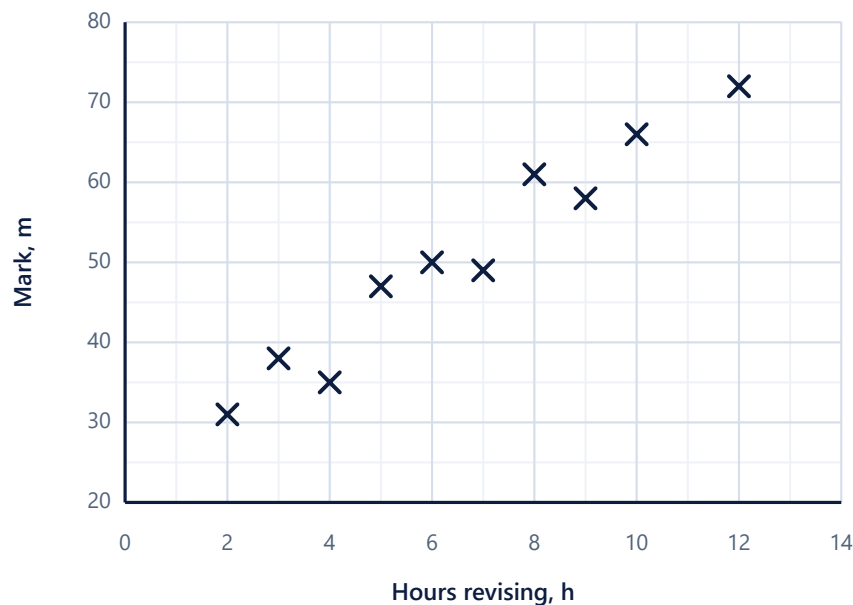




1 Ten students recorded the number of hours, h , they spent revising for a test and their mark, m . The scatter diagram shows the results.

- Describe the correlation between h and m .
- Interpret this correlation in the context of the question.
- State which variable is the explanatory variable.



(3 marks)

a) Points rise from left to right

Positive correlation

b) Say what it means for these students

Students who revised for more hours
tended to score higher marks

c) Which variable affects the other?

Revision time may affect the mark

h is the explanatory variable

2

A café owner records the daily maximum temperature, x °C, and the number of ice creams sold, y , on 20 days in summer. The temperatures ranged from 14 °C to 28 °C.

The equation of the regression line of y on x is

$$y = 12.5 + 3.2x$$

- Give an interpretation of the gradient of the regression line.
- Estimate the number of ice creams sold on a day when the maximum temperature is 20 °C.
- Comment on the reliability of using the line to estimate the number sold on a day when the maximum temperature is 35 °C.

(4 marks)

a) Gradient = change in y for each 1 unit of x

Each 1 °C increase in temperature

goes with about 3.2 more ice creams sold

b) Substitute $x = 20$

$$y = 12.5 + 3.2 \times 20$$

$$y = 12.5 + 64$$

$$y = 76.5$$

About 76 or 77 ice creams

c) Is 35 inside the data range?

Data run from 14 °C to 28 °C

35 °C is outside this range

Extrapolation, so unreliable

3

For each pair of variables, state which is the explanatory variable and which is the response variable, and state whether you would expect positive or negative correlation.

- The amount of fertiliser given to a tomato plant and the mass of tomatoes it produces.
- The outside temperature and the amount of gas a house uses for heating each day.

(4 marks)

Explanatory = the variable that affects the other

a)

Fertiliser affects how many tomatoes grow

Explanatory: amount of fertiliser

Response: mass of tomatoes

More fertiliser, more tomatoes

Positive correlation

b)

Temperature affects how much heating is used

Explanatory: outside temperature

Response: gas used

Colder days, more gas

Negative correlation

4

A council collects data for 15 towns in a region. For each town it records the number of cafés and the number of road traffic accidents last year. The data show strong positive correlation.

A councillor says, "Cafés cause road accidents, so we should stop new cafés opening."

- Explain why the councillor's conclusion is not justified.
- Suggest a variable that could explain the correlation.

(3 marks)

a) Correlation is not causation

The data only show that towns with more cafés tend to have more accidents

This does not mean the cafés cause the accidents

b) Look for something that affects both

Bigger towns have more cafés

and more traffic, so more accidents

The population of the town

The table shows the engine size, x litres, and the fuel economy, y miles per gallon, of eight cars.

Engine size, x (litres)	1.0	1.2	1.4	1.5	1.8	2.0	2.5	3.0
Fuel economy, y (mpg)	56	54	51	50	46	43	37	31

The equation of the regression line of y on x is $y = 68.9 - 12.7x$

- State, giving a reason, which variable is the explanatory variable.
- Give an interpretation of the gradient of the line.
- Estimate the fuel economy of a car with a 1.6 litre engine.
- Explain why the value 68.9 has no practical meaning here.
- Explain why the line should not be used to estimate the engine size of a car that does 40 mpg.

(5 marks)

a) Which affects the other?

x , engine size: it affects fuel economy

b) Gradient = change in y per 1 litre

Each extra litre of engine size

goes with about 12.7 fewer miles per gallon

c) Substitute $x = 1.6$

$$y = 68.9 - 12.7 \times 1.6$$

$$y = 68.9 - 20.32$$

$$y = 48.58$$

48.6 mpg (3 s.f.)

d) 68.9 is the value when $x = 0$

A car cannot have a 0 litre engine

(and 0 is far outside the data)

e) Which way does the line work?

It is the regression line of y on x

so it is only for estimating y from x ,

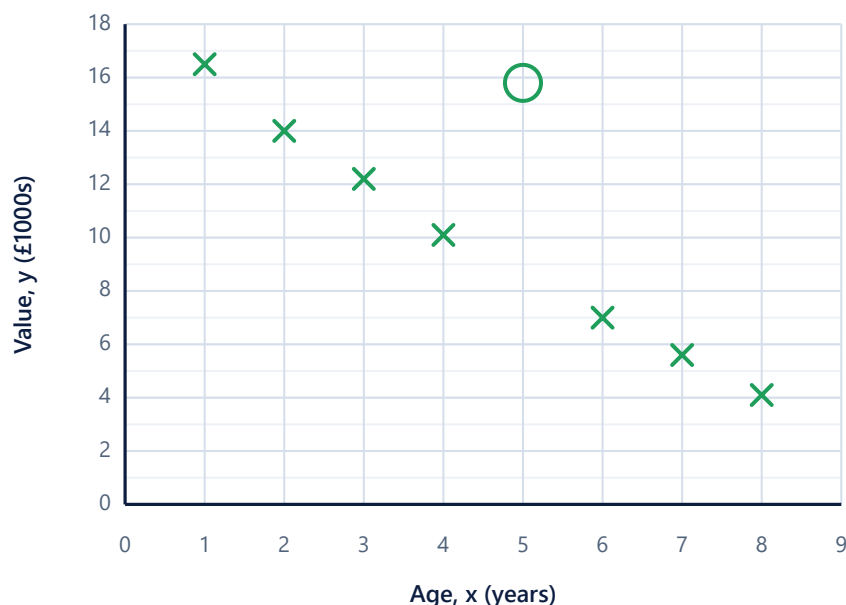
not x from y

6

The table shows the age, x years, and the value, y thousand pounds, of eight cars of the same model sold by a garage.

Age, x (years)	1	2	3	4	5	6	7	8
Value, y (£1000s)	16.5	14.0	12.2	10.1	15.8	7.0	5.6	4.1

- On the grid, draw a scatter diagram of these data.
- One point does not fit the pattern. Write down its coordinates and suggest a reason for it.
- Describe the correlation shown by the other seven points.



(4 marks)

a) Plot each (x, y) with a cross

$(1, 16.5), (2, 14.0), (3, 12.2), (4, 10.1)$

$(5, 15.8), (6, 7.0), (7, 5.6), (8, 4.1)$

b) Which point is away from the rest?

$(5, 15.8)$

This car is worth far more than expected for its age

e.g. very low mileage, or a recording error

c) The other seven points

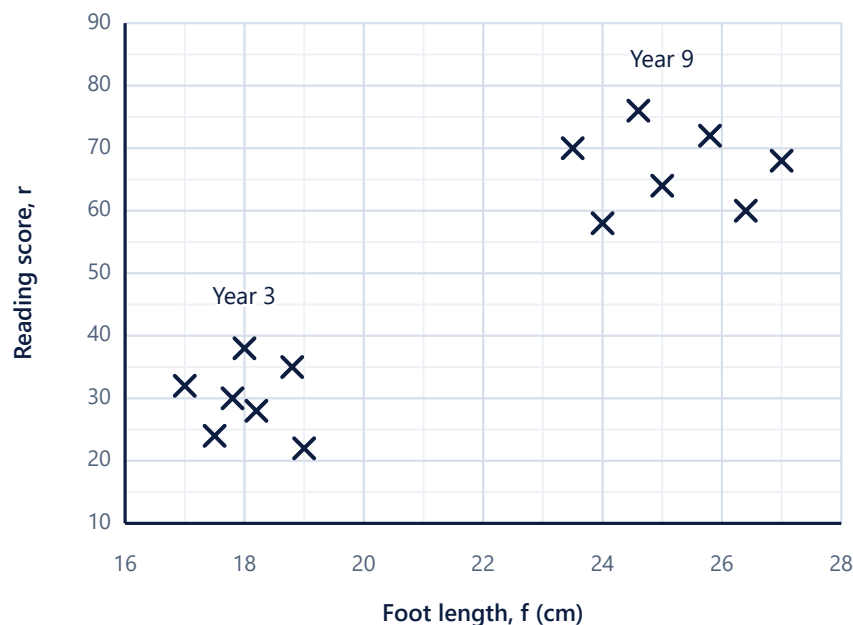
fall from left to right, close to a straight line

Strong negative correlation

7

A researcher measures the foot length, f cm, and the reading score, r , of 14 children: seven from Year 3 and seven from Year 9. The scatter diagram shows the results.

- Describe the correlation shown by all 14 points.
- Kofi says, "Children with longer feet are better readers." With reference to the scatter diagram, comment on Kofi's claim.
- Describe the correlation within each year group.



(4 marks)

a) All 14 points rise from left to right

Positive correlation

b) Look at the shape of the diagram

The points form two separate groups

Year 3: short feet, low scores

Year 9: long feet, high scores

Age causes both, not foot length

Kofi's claim is not supported

c) Look inside each group

No clear pattern in either group

Little or no correlation within each year

8

Lewis uses data from a UK weather station for 31 days in May. He records the daily total sunshine, s hours, and the daily maximum temperature, t °C. The values of s range from 0 to 14.2

The regression line of t on s is $t = 13.4 + 0.62s$

- State the units of the gradient and give an interpretation of the gradient.
- Give an interpretation of the value 13.4
- Estimate the maximum temperature on a day with 8 hours of sunshine.
- The sunshine for one day was recorded as n/a. Explain how Lewis should deal with this day before finding the regression line.

(6 marks)

a) Units of gradient = $\frac{\text{units of } t}{\text{units of } s}$

°C per hour

Each extra hour of sunshine goes with about 0.62 °C higher maximum temperature

b) 13.4 is t when $s = 0$

Estimated maximum temperature on a day with no sunshine is 13.4 °C

c) Substitute $s = 8$

$$t = 13.4 + 0.62 \times 8$$

$$t = 13.4 + 4.96$$

$$t = 18.36$$

18.4 °C (3 s.f.)

d) n/a = not available

There is no sunshine value for that day so remove that day from the data

9

A taxi firm models the cost, £ y , of a journey of x miles by the regression line

$$y = 2.80 + 1.65x$$

- a. Give an interpretation of the values 2.80 and 1.65

The firm decides to record distances, k , in kilometres, where 1 km = 0.621 miles.

- b. Find the equation of the regression line of y on k , giving the gradient to 3 significant figures.
c. Use your equation to estimate the cost of a 12 km journey.

(5 marks)

a) *Intercept and gradient in context*

2.80: a fixed charge of £2.80 before any distance

1.65: each mile costs about £1.65 more

b) *Write x in terms of k*

$$x = 0.621k$$

$$y = 2.80 + 1.65 \times 0.621k$$

$$y = 2.80 + 1.02465k$$

$$y = 2.80 + 1.02k$$

c) *Substitute $k = 12$ (use the unrounded gradient)*

$$y = 2.80 + 1.02465 \times 12$$

$$y = 2.80 + 12.2958$$

$$y = 15.0958$$

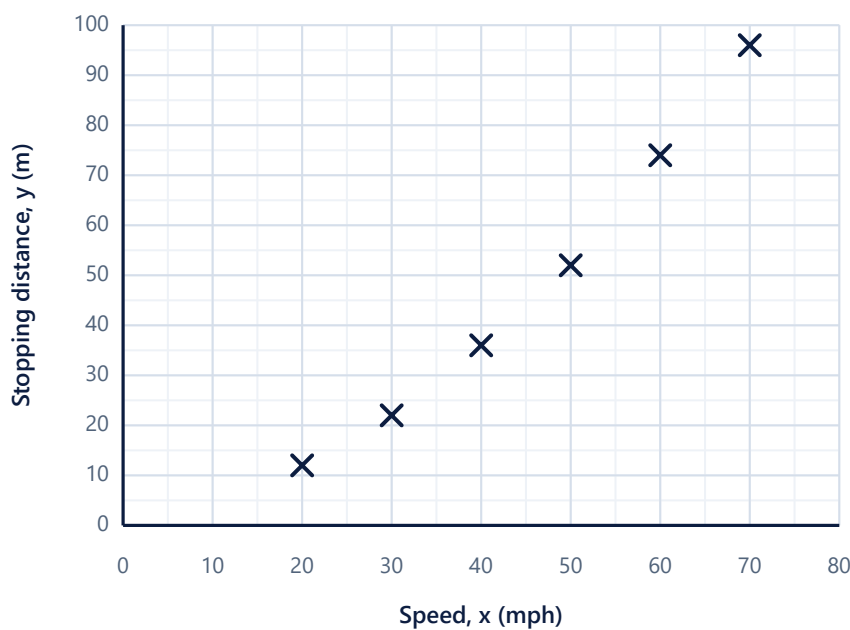
About £15.10

10

The scatter diagram shows the speed, x mph, and the stopping distance, y metres, of a car in six tests. The regression line of y on x is

$$y = -27.4 + 1.69x$$

- Describe the correlation between x and y .
- Give an interpretation of the gradient of the line.
- With reference to the scatter diagram, state, giving a reason, whether a linear regression model is suitable for these data.
- Use the line to estimate the stopping distance at 10 mph, and comment on your answer.



(6 marks)

a) Points rise from left to right

Strong positive correlation

b) Gradient = change in y per 1 mph

Each extra 1 mph adds about 1.69 m to the stopping distance

c) Look at the shape of the points

The points lie on a curve that gets steeper
A straight line is too high in the middle and too low at each end

Not suitable: the relationship is curved

d) Substitute $x = 10$

$$y = -27.4 + 1.69 \times 10$$

$$y = -27.4 + 16.9$$

$$y = -10.5$$

A stopping distance cannot be negative
10 mph is outside the data: extrapolation

11

An ice cream seller works at two sites. She models her daily takings, £ y , against the daily maximum temperature, x °C.

Site	Regression line of y on x	Temperatures in the data
Beach	$y = -120 + 22x$	15 °C to 30 °C
Park	$y = 40 + 9x$	12 °C to 28 °C

- Give an interpretation of the value 22 in the beach model.
- Find the temperature at which the two models predict the same takings.
- Comment on the reliability of your answer to part (b).
- The forecast maximum temperature for tomorrow is 26 °C. Find how much more the models predict she would take at the beach than at the park.

(7 marks)

a) 22 is the gradient of the beach line

Each 1 °C rise in temperature

goes with about £22 more takings at the beach

b) Set the two predictions equal

$$-120 + 22x = 40 + 9x$$

$$13x = 160$$

$$x = \frac{160}{13}$$

$$x = 12.30\dots$$

$$x = 12.3^\circ\text{C}$$

c) Is 12.3 inside both data ranges?

Park: 12 to 28, so yes

Beach: 15 to 30, so no

Extrapolation for the beach, so unreliable

d) Substitute $x = 26$ in both

$$\text{Beach: } y = -120 + 22 \times 26 = 452$$

$$\text{Park: } y = 40 + 9 \times 26 = 274$$

$$\text{Difference} = 452 - 274$$

£178 more at the beach

12

A school nurse records the height, h cm, and the mass, m kg, of 25 fourteen-year-olds. The heights range from 148 cm to 182 cm. She codes the data using

$$x = \frac{h - 150}{5} \quad \text{and} \quad y = m - 40$$

The regression line of y on x is $y = 4.2 + 0.85x$

- Find the equation of the regression line of m on h in the form $m = a + bh$
- Give an interpretation of the value of b .
- Estimate the mass of a fourteen-year-old who is 170 cm tall.
- The nurse uses the line to estimate the mass of an adult who is 195 cm tall. Give two reasons why this estimate may be unreliable.

(8 marks)

a) Substitute both codes into the line

$$m - 40 = 4.2 + 0.85 \times \frac{h - 150}{5}$$

$$m - 40 = 4.2 + 0.17(h - 150)$$

$$m - 40 = 4.2 + 0.17h - 25.5$$

$$m = 40 + 4.2 - 25.5 + 0.17h$$

$$m = 18.7 + 0.17h$$

b) b is the gradient

Each extra 1 cm of height goes with about 0.17 kg more mass

c) Substitute $h = 170$

$$m = 18.7 + 0.17 \times 170$$

$$m = 18.7 + 28.9$$

$$m = 47.6 \text{ kg}$$

d) Two reasons

195 cm is outside the data range (148 to 182 cm), so this is extrapolation

The data are for fourteen-year-olds, so the model may not apply to adults