GCSE Mathematics

Scatter Graphs

Questions

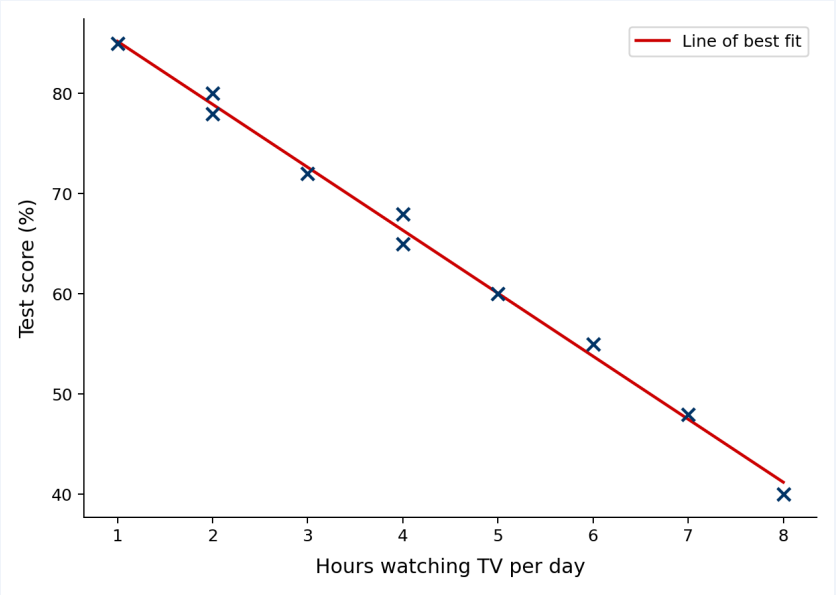
Pearson Edexcel GCSE & iGCSE Mathematics

Section A — Foundation

Worked Examples

[Fluency]

Describe the correlation shown in the scatter graph.



Negative correlation — as TV hours increase, test scores decrease.

[Reasoning]

Draw a line of best fit and use it to predict the test score for 3.5 hours of TV.

Read from line at x=3.5: test score ≈ **70%**

[Problem Solving]

Explain why it would be unreliable to use this graph to predict the score for 15 hours of TV.

15 hours is **outside the range** of the data (extrapolation) — the relationship may not continue.

[Fluency]

1.

Describe the type of correlation shown in each scatter graph description:

(a) As height increases, weight increases. (1)

(b) As temperature increases, coat sales decrease. (1)

(c) Shoe size and exam score. (1)

(3 marks)

[Fluency]

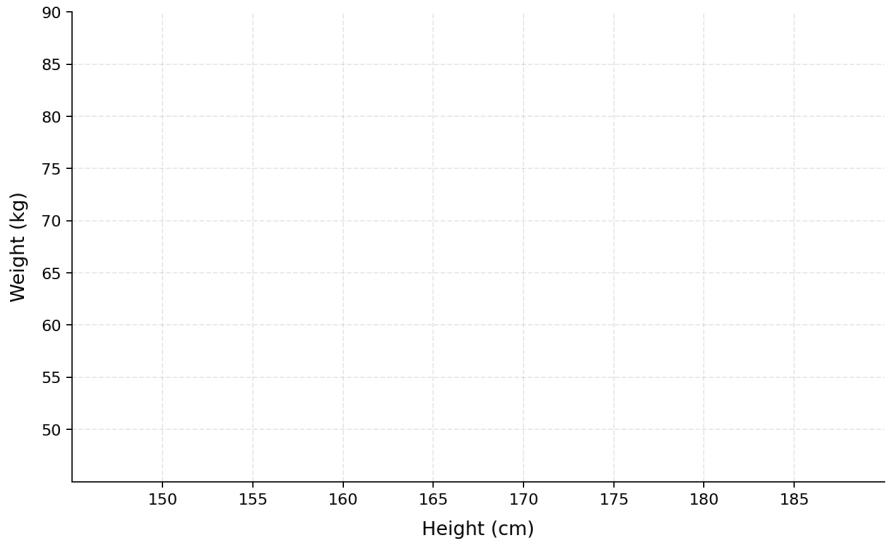
2.

Height (cm)	150	155	160	165	170	175	180	185
Weight (kg)	52	55	60	65	68	72	78	82

(a) Plot the data on a scatter graph. (2)

(b) Describe the correlation. (1)

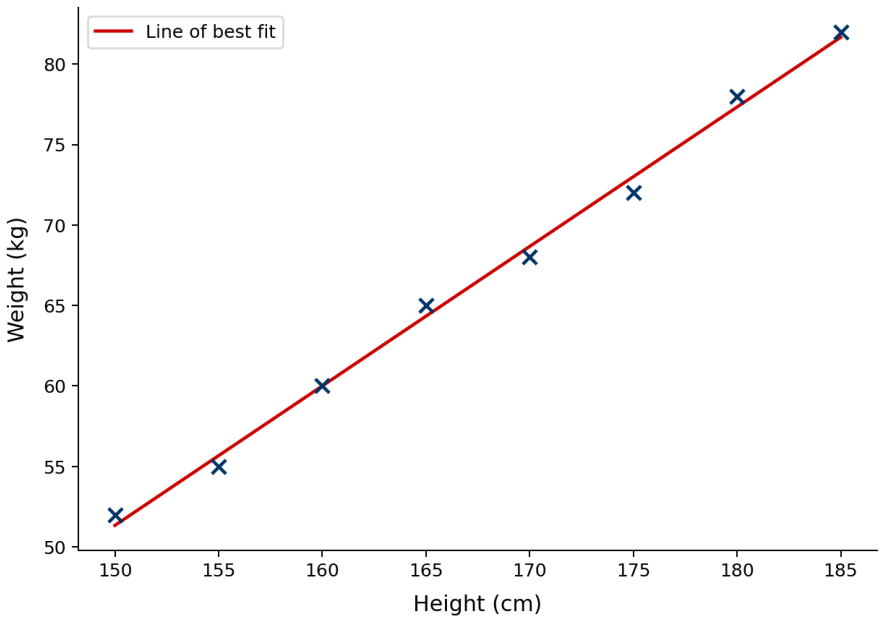
(c) Draw a line of best fit. (1)



(4 marks)

[Fluency]

3.



Use the line of best fit to estimate the weight of a person 172 cm tall.

(2 marks)

[Reasoning]

4.

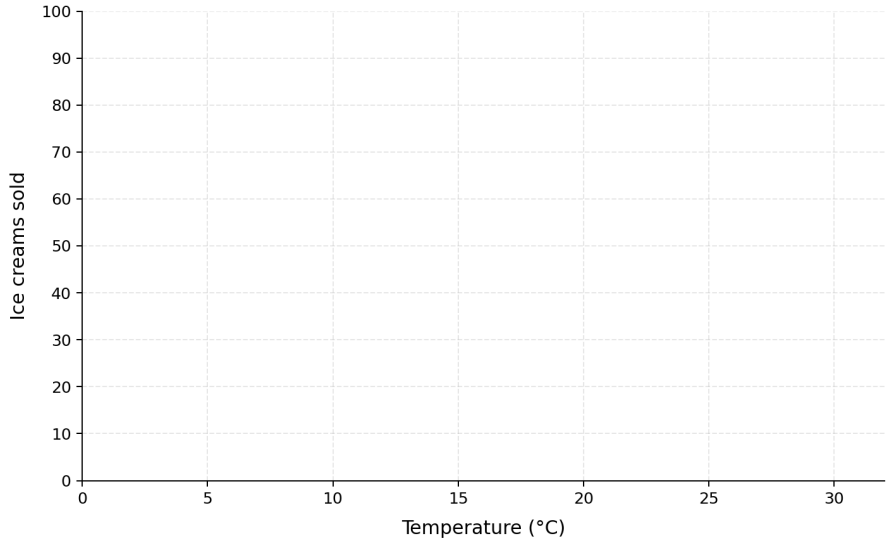
Temperature (°C)	5	8	10	12	15	18	20	22	25	28
Ice creams sold	20	28	35	42	55	68	72	80	88	95

(a) Plot the scatter graph. (2)

(b) Describe the correlation. (1)

(c) Draw a line of best fit. (1)

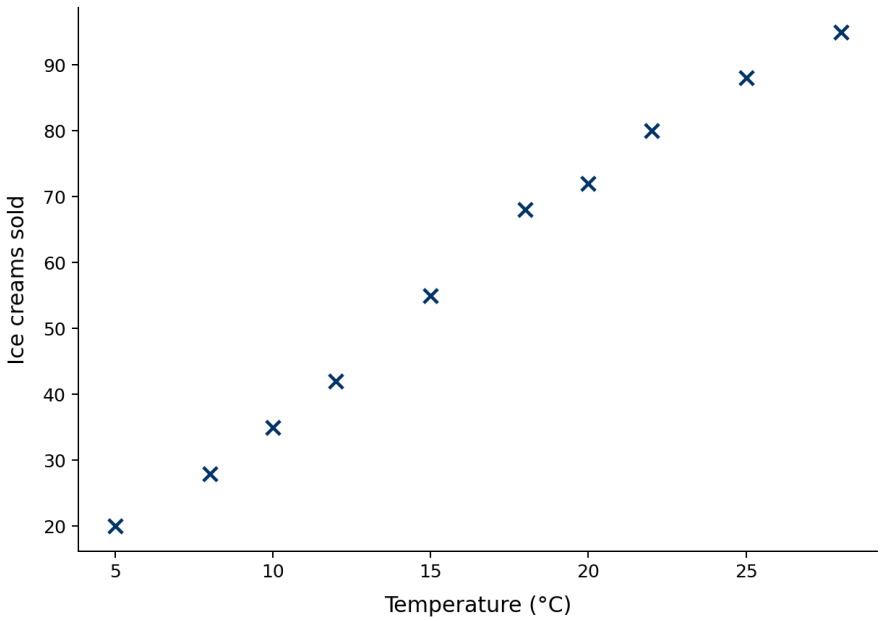
(d) Estimate the number of ice creams sold at 16°C. (1)



(5 marks)

[Reasoning]

5.



(a) Draw a line of best fit. (1)

(b) Use your line to estimate ice cream sales at 30°C. (1)

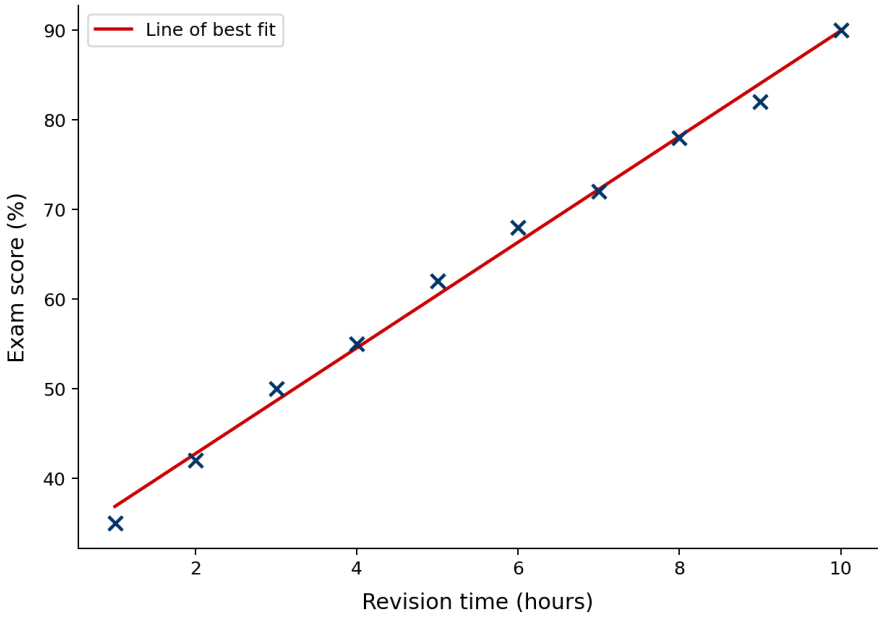
(c) Is this estimate reliable? Give a reason. (2)

(4 marks)

[Reasoning]

6.

The scatter graph shows revision time (hours) vs exam score (%).



(a) Describe the correlation. (1)

(b) One student revised for 5 hours but scored 40%. Circle this point and suggest a reason for the anomaly. (2)

(c) Use the line to estimate the score for 6.5 hours of revision. (1)

(4 marks)

[Problem Solving]

7.

Explain the difference between **interpolation** and **extrapolation**.
Which is more reliable, and why?

(3 marks)

[Problem Solving]

8.

A student says: “The scatter graph shows that watching more TV *causes* lower test scores.”
Explain why this statement may not be correct.

(2 marks)

Section B — Higher

Worked Examples

[Fluency]

Calculate the equation of the line of best fit for the height/weight data.

$w = 0.867h - 78.67$

Gradient ≈ 0.867 kg/cm y-intercept ≈ –78.67

[Reasoning]

The PMCC is r = 0.998. What does this tell us?

r is very close to 1 → **very strong positive correlation**.

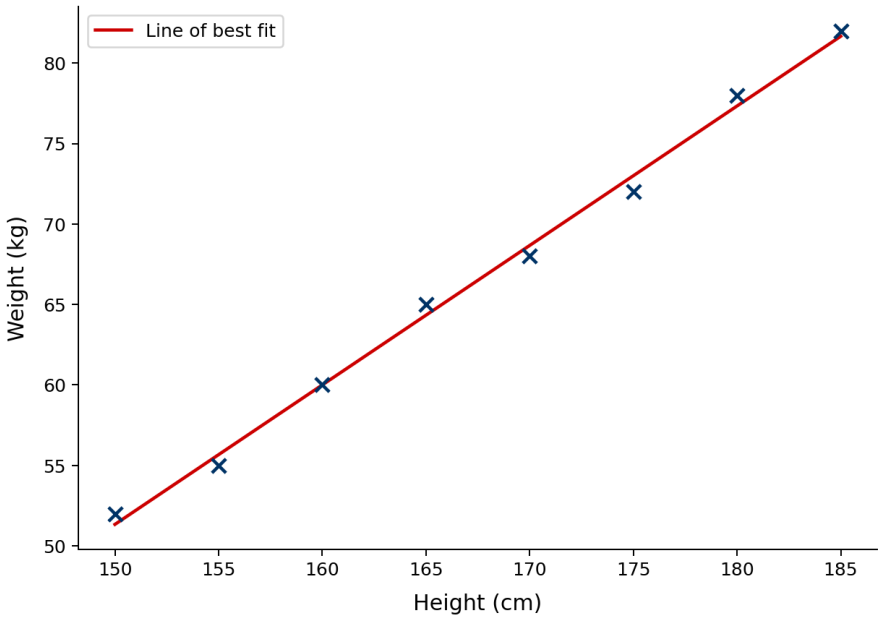
[Problem Solving]

Distinguish between correlation and causation.

Correlation means two variables are related. Causation means one causes the other. Correlation does not imply causation — a third factor (lurking variable) may be responsible.

[Fluency]

1.



(a) Find the equation of the line of best fit in the form w = mh + c. (3)

(b) Use your equation to predict the weight of a person 168 cm tall. (2)

(c) Comment on the reliability of predicting for h = 210 cm. (1)

(6 marks)

[Fluency]

2.

The PMCC for a dataset is r = –0.92.

(a) Describe the correlation. (1)

(b) What does this value suggest about the relationship between the two variables? (2)

(3 marks)

[Reasoning]

3.

A scatter graph of shoe size vs spelling test score shows r = 0.85 for primary school children.

A student concludes: “Having bigger feet improves spelling.”

Explain the flaw in this reasoning and suggest an alternative explanation.

(3 marks)

[Reasoning]

4.

x	2	4	5	7	8	10	12	14
y	14	11	10	8	6	5	3	1

(a) Plot the scatter graph and draw a line of best fit. (3)

(b) Find the equation of the line of best fit. (3)

(c) Estimate y when x = 9. (1)

(7 marks)

[Reasoning]

5.

Two variables have a PMCC of r = 0.04.

(a) Describe the correlation. (1)

(b) Is it appropriate to draw a line of best fit? Explain. (2)

(3 marks)

[Problem Solving]

6.

The table shows data on advertising spend (£000s) and monthly sales (£000s):

Advert (£k)	2	3	5	5	7	8	10
Sales (£k)	18	22	28	31	35	40	48

(a) Plot the data and draw a line of best fit. (3)

(b) Find the equation of the line of best fit. (3)

(c) Predict sales when advertising spend is £6,000. (1)

(d) Give one reason why your prediction may not be accurate. (1)

(8 marks)