6.7 Scatter Graphs and Correlation: 30-minute Pack

Calculators not allowed · about 30 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 A scatter graph is drawn from these points: (3, 15), (4, 17), (5, 8), (7, 25), (10, 36), (15, 49), (19, 62). The line of best fit for the other points is y = 3x + 5.
One point is an outlier. Write down the coordinates of the outlier.

(2)

---

2 The scatter graph plots shoe size (x) and reading score (y) for 8 cases. The points are: (1, 69), (6, 49), (8, 47), (9, 42), (10, 42), (11, 38), (12, 31), (15, 21).
Describe the correlation shown by the points.

(2)

---

3 A scatter graph is drawn from these points: (2, 17), (3, 36), (9, 52), (10, 55), (14, 76), (15, 80), (17, 92). The line of best fit for the other points is y = 5x + 6.
One point is an outlier. Write down the coordinates of the outlier.

(2)

---

4 The scatter graph plots shoe size (x) and reading score (y) for 7 cases. The points are: (1, 12), (3, 4), (4, 4), (5, 4), (7, 20), (10, 18), (12, 5).
Describe the correlation shown by the points.

(2)

© Revision Library | revisionlibrary.co.uk

5 Given points (2, 3), (4, 7), (6, 11), explain whether these points lie on a straight line and justify your answer.

(2)

---

6 A scatter of body mass (kg) against vertical jump height (cm) shows a weak negative correlation. Give two possible real-world explanations for this correlation (not calculations).

(2)

---

7 A student suggests that 'correlation implies causation'. Give one brief reason why this statement may be incorrect when looking at scatter graphs.

(2)

---

8 A set of points has approximate line of best fit $y = 0.5x + 2$. A pupil reads off the graph and finds the point (6, 9) lies near the line but slightly above it. Calculate the residual for the point (6, 9) using the line $y = 0.5x + 2$ (residual = actual y - predicted y).

(2)

---

9 A pupil draws two different possible lines of best fit on the same scatter. Give one reason why different people might draw different lines of best fit and one way to make the estimate more reliable.

(2)

© Revision Library | revisionlibrary.co.uk

10 A scatter graph uses two axes. State what each axis represents and give a sensible label for the horizontal axis for a study of hours studied (x) and test score (y).

(a) What does the horizontal axis represent on a scatter graph? **(1)**

(b) Give a sensible label for the horizontal axis when comparing hours studied and test score. **(1)**

---

11 A line of best fit for temperature (x) and ice creams sold (y) has equation $y = 6x + 13$. Write down what the gradient 6 means in this context.

(2)

---

12 A line of best fit for temperature (x) and ice creams sold (y) has equation $y = 2x + 29$. Use the line to estimate y when $x = 7$.

(2)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 6.7 Scatter Graphs and Correlation: 30-minute Pack

Question 1

compares the points with the line of best fit — **M1**

(5, 8) cao — **A1**

Answer: **(5, 8)**

Question 2

describes the trend (as x increases, y ...) in words or the type of correlation — **M1**

negative correlation cao — **A1**

Answer: **negative correlation**

Question 3

compares the points with the line of best fit — **M1**

(3, 36) cao — **A1**

Answer: **(3, 36)**

Question 4

describes the trend (as x increases, y ...) in words or the type of correlation — **M1**

no correlation cao — **A1**

Answer: **no correlation**

Question 5

check equal differences in x and corresponding equal differences in y or calculate gradient between pairs — **M1**

state yes they lie on a straight line because gradient is 2 between each consecutive pair (cao) — **A1**

Answer: **Yes. Gradient between pairs = (7-3)/(4-2)=2 and (11-7)/(6-4)=2 so points lie on the straight line y = 2x - 1**

Question 6

one plausible explanation linking higher mass to lower jump height or measurement bias — **M1**

second plausible explanation such as difference in training levels or age affecting results — **A1**

Answer: **Examples: heavier participants may jump less because extra mass makes jumping harder; also those who train more (lighter or fitter) may jump higher, causing the weak negative pattern**

Question 7

identify that correlation may be due to a third factor or coincidence — **M1**

state that without controlled experiment you cannot prove cause and effect cao — **A1**

Answer: **Correlation may be caused by a third variable or by coincidence; only controlled experiments can show causation**

Question 8

substitute x = 6 to get predicted y = 0.5(6) + 2 = 5 — **M1**

residual = 9 - 5 = 4 cao — **A1**

Answer: **4**

Question 9

reason: subjective judgement on where most points cluster or whether to balance points above and below the line — **M1**

way to improve reliability: use more data points or use statistical method (least squares) or take average of repeated estimates — **A1**

Answer: **Reason: different subjective judgements about the best balance of points; Improve: collect more data or use a mathematical fitting method (e.g. least squares)**

Question 10

(a) independent variable or the variable that is changed (cao) — **B1**

(a) Answer: **Independent variable (the variable that is changed)**

(b) label such as 'Hours studied' or 'Time spent studying (hours)' (cao) — **B1**

(b) Answer: **Hours studied**

Question 11

refers to the change in y for each 1 unit increase in x — **M1**

y increases by 6 for each extra unit of x — **A1**

Answer: **ice creams sold increases by 6 for each extra unit of temperature**

Question 12

substitutes x = 7 — **M1**

43 cao — **A1**

Answer: **43**

© Revision Library | revisionlibrary.co.uk