



KS3.M-S7 Scatter Graphs and Correlation

AQA KS3.M-S7 · Calculators not allowed · about 50 minutes

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

1 Plot the following points on graph paper: (1, 2), (3, 4), (5, 6).

(a) Plot the point (1, 2). (1)

(b) Plot the point (3, 4). (1)

(c) Plot the point (5, 6). (1)

2 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)

3 Here are five pairs of data showing hours of revision and exam score: (0, 45), (1, 50), (2, 55), (3, 60), (4, 65). Describe the correlation and say if it is strong, moderate or weak.

(2)

4 For each set of points, state whether the correlation is positive, negative or no correlation.

(a) Points: (1, 1), (2, 2), (3, 3) (1)

(b) Points: (1, 5), (2, 3), (3, 1) (1)

(c) Points: (1, 3), (2, 2), (3, 3) (1)

5 A small company recorded advertising spend (x , in hundreds of pounds) and number of new customers (y): (1, 4), (2, 6), (3, 9), (4, 11). Which point is the best candidate for an outlier? Explain briefly.

(2)

6 Use the pairs: (2, 5), (4, 9), (6, 13). Draw (or imagine) the scatter and write down the equation of the straight line that fits these points exactly in the form $y = mx + c$.

(2)

7 A line of best fit for data is $y = 3x + 2$. Estimate y when $x = 4$.

(2)

8 Explain what the gradient of a line of best fit tells you in the context of a scatter graph where x is time spent practising and y is number of free throws scored.

(2)

- 9** A teacher draws a rough line of best fit through points and uses it to estimate. The line passes through (0, 10) and (5, 25). Calculate the gradient of this line and explain what it means for the data.

(2)

-
- 10** Use the data: (1, 2), (2, 3), (3, 5), (4, 7). Calculate the mean of the x-values and the mean of the y-values. Give your answers as exact values.

(2)

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

(2)

-
- 12** 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)

-
- 13** A line of best fit drawn through data approximately passes through (1, 4) and (5, 16). Use these points to write an equation for the line in the form $y = mx + c$ and then estimate y at $x = 3$.

(2)

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

(2)

- 15** 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)

- 16** Stretch: A line of best fit drawn through a scatter is $y = 4x - 6$. A point at $x = 2$ has $y = 3$. Calculate how far vertically this point is from the line (distance between actual y and line y), and say whether the point lies above or below the line.

(3)

- 17** Stretch: A researcher plots number of weekly practice hours (x) against penalty success rate (%) (y) and finds a strong positive correlation. She fits the line $y = 6x + 20$. Using this line, estimate how many hours of practice are needed to reach a predicted success rate of 80%. Show your working.

(3)

18 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)

Mark scheme · KS3.M-S7 Scatter Graphs and Correlation

Question 1

- (a) **B1** point plotted at $x=1$, $y=2$ (cao)
- (a) Answer: **(1, 2)**
- (b) **B1** point plotted at $x=3$, $y=4$ (cao)
- (b) Answer: **(3, 4)**
- (c) **B1** point plotted at $x=5$, $y=6$ (cao)
- (c) Answer: **(5, 6)**

Question 2

- (a) **B1** independent variable or the variable that is changed (cao)
- (a) Answer: **Independent variable (the variable that is changed)**
- (b) **B1** label such as 'Hours studied' or 'Time spent studying (hours)' (cao)
- (b) Answer: **Hours studied**

Question 3

- **M1** identify positive correlation (scores increase as hours increase) or equivalent
- **A1** describe as strong (points lie close to a straight increasing trend) cao
- Answer: **Positive correlation, strong**

Question 4

- (a) **B1** positive correlation cao
- (a) Answer: **Positive correlation**
- (b) **B1** negative correlation cao
- (b) Answer: **Negative correlation**
- (c) **B1** no correlation or very weak correlation cao
- (c) Answer: **No correlation**

Question 5

- **M1** identify (1, 4) or (4, 11) as less consistent with trend or state (1,4) is lower than expected or (4,11) slightly higher
- **A1** brief explanation: it does not follow the overall increasing pattern as closely as the others cao
- Answer: **(1, 4) is a likely outlier; it is lower than the trend suggested by the other points**

Question 6

- **M1** find gradient $m = (9-5)/(4-2) = 2$ or equivalent
- **A1** $y = 2x + 1$ cao
- Answer: **$y = 2x + 1$**

Question 7

- **M1** substitute $x = 4$ into expression $y = 3x + 2$ or equivalent
- **A1** $y = 14$ cao
- Answer: **14**

Question 8

- **M1** mention relationship between change in y and change in x (rise over run) or rate of change
- **A1** contextual statement: how many extra free throws are scored per unit of practice time (cao)
- Answer: **Gradient = increase in free throws per unit of practice time; it tells how many more free throws are expected for each extra unit of practice**

Question 9

- **M1** calculate gradient = $(25-10)/(5-0) = 3$
- **A1** interpretation: y increases by 3 for each increase of 1 in x, cao
- Answer: **Gradient = 3; it means y increases by 3 for each 1 increase in x**

Question 10

- **M1** calculate mean $x = (1+2+3+4)/4 = 10/4 = 2.5$ or $5/2$
- **A1** calculate mean $y = (2+3+5+7)/4 = 17/4 = 4.25$ or $17/4$ cao
- Answer: **mean x = 5/2; mean y = 17/4**

Question 11

- **M1** check equal differences in x and corresponding equal differences in y or calculate gradient between pairs
- **A1** state yes they lie on a straight line because gradient is 2 between each consecutive pair (cao)
- 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 12

- **M1** one plausible explanation linking higher mass to lower jump height or measurement bias
- **A1** second plausible explanation such as difference in training levels or age affecting results
- 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 13

- **M1** calculate gradient $m = (16-4)/(5-1) = 12/4 = 3$
- **A1** $y = 3x + 1$ and estimate $y(3) = 10$ cao
- Answer: **$y = 3x + 1$; estimate at $x=3$ is 10**

Question 14

- **M1** identify that correlation may be due to a third factor or coincidence
- **A1** state that without controlled experiment you cannot prove cause and effect cao
- Answer: **Correlation may be caused by a third variable or by coincidence; only controlled experiments can show causation**

Question 15

- **M1** substitute $x = 6$ to get predicted $y = 0.5(6) + 2 = 5$
- **A1** residual = $9 - 5 = 4$ cao
- Answer: **4**

Question 16

- **M1** substitute $x = 2$ to find line $y = 4(2) - 6 = 2$
- **M1** calculate vertical distance = $|3 - 2| = 1$
- **A1** state point is 1 above the line (actual y greater than line y) cao
- Answer: **1 above the line**

Question 17

- **M1** set up equation $6x + 20 = 80$ or equivalent
- **M1** solve $6x = 60$ to get $x = 10$ (method shown)
- **A1** conclude 10 hours (cao)
- Answer: **10 hours**

Question 18

- **M1** reason: subjective judgement on where most points cluster or whether to balance points above and below the line
- **A1** way to improve reliability: use more data points or use statistical method (least squares) or take average of repeated estimates
- 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)**