4.20 Scatter Graphs: 1-hour Pack

EDEXCEL 1MA1 · Calculator allowed · about 60 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 For each situation below, state whether you would expect positive correlation, negative correlation or no correlation between the two quantities.

(a) The number of pages in a book and the weight of the book. **(1)**

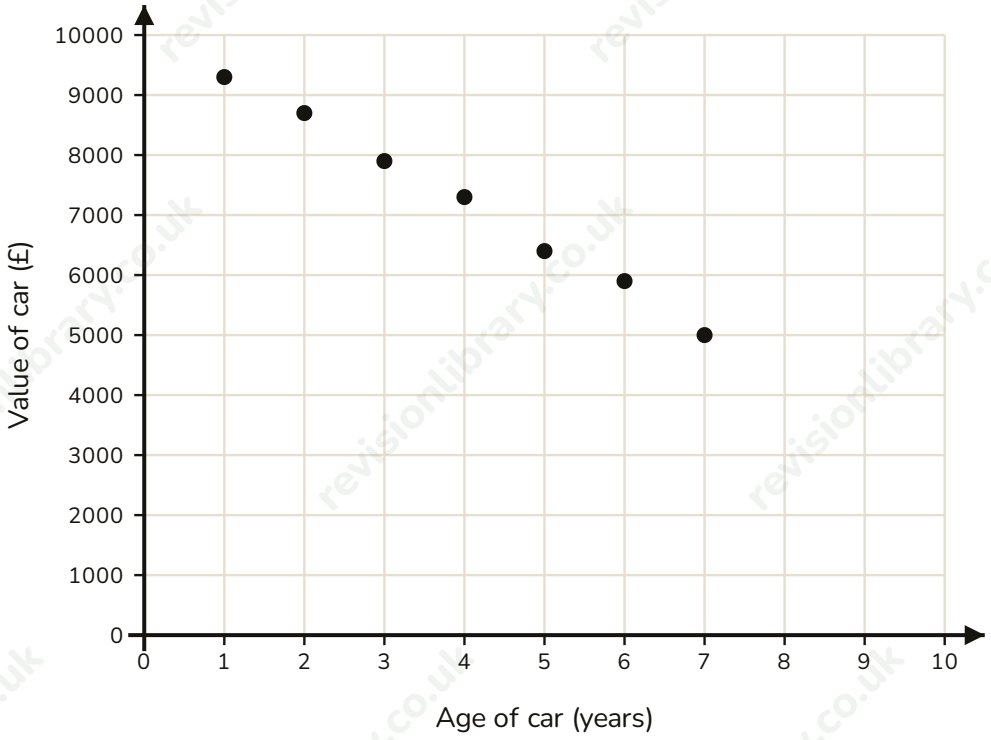
(b) The age of a car and its value in £. **(1)**

(c) A person's shoe size and their favourite colour. **(1)**

(Total for Question 1 is 3 marks)

© Revision Library | revisionlibrary.co.uk

2 A garage records the age and value of used cars of the same model. Age in years (x): 1, 2, 3, 4, 5, 6, 7, 8, 9. Value in pounds (y): 9300, 8700, 7900, 7300, 6400, 5900, 5000, 4500, 3700. The points for ages 1 to 7 have already been plotted on the grid.



(a) Plot the remaining two points from the table on the grid. **(2)**

(b) A line of best fit has been drawn passing through (0, 10000) and (10, 3000). Use this line to estimate the value of a car that is 6 years old. **(2)**

(c) Explain why it would not be appropriate to use this line of best fit to predict the value of the car when it is 25 years old. **(2)**

(d) State what the gradient of the line of best fit represents in this context. **(1)**

(Total for Question 2 is 7 marks)

© Revision Library | revisionlibrary.co.uk

3 Nine students record how many hours they revised and their test score (%). Hours revised (x): 1, 2, 3, 4, 5, 6, 6, 7, 8. Test score % (y): 40, 45, 52, 58, 20, 68, 72, 78, 82.

(a) Write down the coordinates of the outlier. **(1)**

.....

(b) Suggest a possible reason for this outlier. **(1)**

.....

(Total for Question 3 is 2 marks)

---

4 A small company records advertising spend per week (x in hundreds of pounds) and number of online orders (y) for 6 weeks: (2, 30), (3, 40), (4, 45), (5, 50), (3.5, 42), (6, 58).

(a) Draw or imagine a line of best fit. Describe the correlation in one sentence (positive/negative and strength). **(1)**

(b) Estimate the number of orders when x = 4.5 (i.e. £450). **(3)**

(c) One week the spend was 1 (i.e. £100) and orders were 20. Is this point an outlier compared with the trend? Give a short reason. **(2)**

.....

.....

(Total for Question 4 is 6 marks)

---

5 A scientist plots the reaction time (in seconds) of runners against their age (in years), for ages 10 to 80. The points curve upwards steeply for ages under 20 and over 60, but are roughly flat for ages 20 to 60. Explain why it would not be appropriate to draw a single straight line of best fit for this data.

.....

.....

.....

(Total for Question 5 is 2 marks)

© Revision Library | revisionlibrary.co.uk

6 A teacher recorded the number of hours 12 students spent revising and their score (%) in a maths test. The scatter graph shows strong positive correlation, and the line of best fit passes through (0, 30) and (10, 90).

(a) Use the line of best fit to estimate the score of a student who revised for 12 hours, and explain why this estimate may not be reliable. **(3)**

(b) The teacher claims 'revising for longer causes a higher test score'. Comment on whether the scatter graph supports this claim. **(2)**

(Total for Question 6 is 5 marks)

---

7 Eight players record their number of hours of sleep the night before a computer game tournament (x) and their reaction time in milliseconds in a test round (y); a lower reaction time means a faster response.

Hours of sleep (x): 4, 5, 6, 7, 8, 9, 10, 11

Reaction time in ms (y): 340, 310, 290, 265, 105, 230, 200, 180

(a) Write down the coordinates of the outlier. **(1)**

.....

(b) Suggest a possible reason for this outlier. **(1)**

(Total for Question 7 is 2 marks)

© Revision Library | revisionlibrary.co.uk

8 Six children record the number of pages they read each day and the number of minutes they spend reading that day.

Pages read per day (x): 3, 5, 6, 8, 9, 11

Minutes spent reading (y): 15, 25, 30, 42, 47, 57

(a) Calculate the mean number of pages read per day. **(1)**

.....

(b) Calculate the mean number of minutes spent reading. **(1)**

.....

(c) Write down the coordinates of the point that must lie on the line of best fit. **(1)**

.....

(Total for Question 8 is 3 marks)

© Revision Library | revisionlibrary.co.uk

9 Graph A shows outdoor temperature plotted against sun cream sales at a chemist's, with points lying close to a straight line. Graph B shows shoe size plotted against spelling test score for the same group of people, with points scattered across the grid with no clear pattern.

(a) State which graph shows the stronger correlation. **(1)**

(b) Give a reason for your answer. **(1)**

.....

(Total for Question 9 is 2 marks)

---

10 For each statement, write True or False.

(a) A strong positive correlation between two variables always means that increasing one variable directly causes the other to increase. **(1)**

.....

(b) The point given by the mean of the x-values and the mean of the y-values always lies on the line of best fit. **(1)**

.....

(c) A scatter graph in which the points are spread randomly across the grid, with no visible pattern, shows no correlation. **(1)**

.....

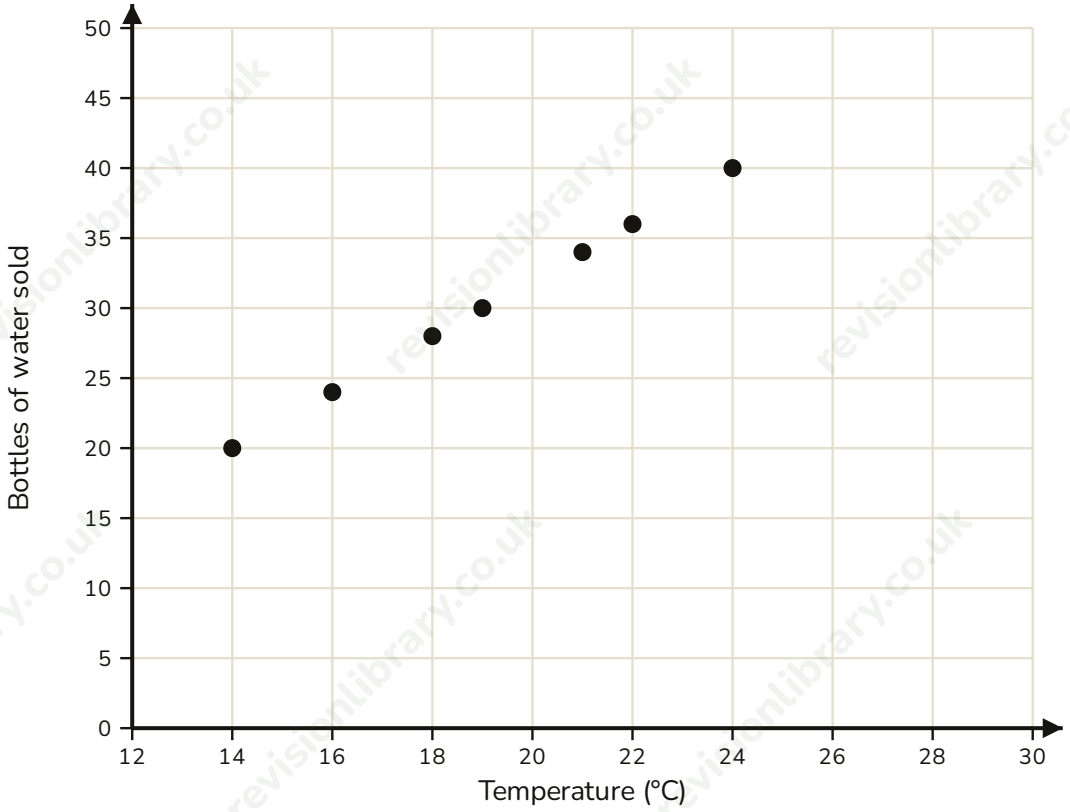
(d) A strong correlation between the number of people visiting a beach and the number of ice creams sold proves that buying ice cream makes people want to visit the beach. **(1)**

.....

(Total for Question 10 is 4 marks)

© Revision Library | revisionlibrary.co.uk

11 A market stall records the outdoor temperature (in °C) and the number of bottles of water it sells on 9 days. The first 7 days have already been plotted on the grid.
Temperature in °C (x): 14, 16, 18, 19, 21, 22, 24
Bottles of water sold (y): 20, 24, 28, 30, 34, 36, 40
On the remaining two days, the temperature was 26°C with 44 bottles sold, and 28°C with 48 bottles sold.



(a) Plot the remaining two points on the grid. **(2)**

(b) Draw a line of best fit on the scatter graph. **(1)**

(c) Use your line of best fit to estimate the number of bottles of water sold when the temperature is 23°C. **(1)**

(Total for Question 11 is 4 marks)

© Revision Library | revisionlibrary.co.uk

12 A teacher records, for 8 pupils, the number of hours each pupil spent using a revision app in the week before a test (x) and their test score out of 50 (y).
Hours on app (x): 1, 2, 3, 4, 5, 6, 7, 8
Test score out of 50 (y): 22, 25, 8, 30, 34, 38, 41, 45

(a) Write down the coordinates of the outlier. **(1)**

.....

(b) The teacher removes this outlier before drawing a line of best fit for the remaining 7 pupils. Explain why this is likely to make the line of best fit a more useful model of the relationship between hours on the app and test score. **(2)**

(c) State one reason why it would not be appropriate to use this line of best fit to predict the test score of a pupil who spends 20 hours a week on the app. **(1)**

(Total for Question 12 is 4 marks)

13 A scientist measures the wing length, in mm, and the flying speed, in mph, of 12 swifts. The wing lengths recorded were between 165 mm and 210 mm.

(a) A birdwatcher uses the line of best fit to estimate the flying speed of a swift with a wing length of 195 mm. State whether this estimate is likely to be reliable, giving a reason. **(2)**

(b) The birdwatcher then uses the same line of best fit to estimate the flying speed of a swift with a wing length of 260 mm. State whether this estimate is likely to be reliable, giving a reason. **(2)**

(Total for Question 13 is 4 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 4.20 Scatter Graphs: 1-hour Pack

Question 1

(a) positive correlation **B1**

(a) Answer: Positive correlation

(b) negative correlation **B1**

(b) Answer: Negative correlation

(c) no correlation **B1**

(c) Answer: No correlation

Question 2

(a) point (8, 4500) plotted correctly **B1**

(a) point (9, 3700) plotted correctly **B1**

(a) Answer: (8, 4500) and (9, 3700)

(b) correct method, e.g. gradient = (3000-10000)/10 = -700, then 10000 - 700 x 6 **M1**

(b) £5800 cao **A1**

(b) Answer: £5800

(c) 25 years is outside the range of the data collected (up to 9 years), so this is extrapolation **B1**

(c) at 25 years the line would give a negative value (10000 - 700x25 = -7500), which is impossible, showing the linear trend cannot continue **B1**

(c) Answer: 25 years is far outside the data range, and substituting gives a negative value (-£7500), which is impossible, so the line of best fit cannot be used here.

(d) the value of the car decreases by about £700 for each extra year of age **B1**

(d) Answer: The value of the car decreases by approximately £700 for each additional year of age.

Question 3

(a) (5, 20) cao **B1**

(a) Answer: (5, 20)

(b) any sensible reason, e.g. the student may have been unwell on the day of the test, or the score may have been recorded incorrectly (oe) **B1**

(b) Answer: The student may have been unwell or unable to prepare properly on that occasion.

Question 4

(a) positive correlation identified **B1**

(a) Answer: Positive correlation, moderately strong.

(b) use interpolation between (4,45) and (5,50) or use line of best fit to find y at x = 4.5 **M1**

(b) calculate midpoint: 45 + 0.5*(50-45) = 47.5 or equivalent **M1**

(b) 47.5 orders cao (allow 47 to 48 awrt) **A1**

(b) Answer: Approximately 47.5 orders

(c) identify it as an outlier or not **M1**

(c) reason: it is below the trend (fewer orders than expected for that spend) or lies far from the line of best fit cao **A1**

(c) Answer: Yes, it is an outlier, because it lies well below the trend/line of best fit for the other points.

Question 5

the relationship is not linear (it curves / changes shape across the range) **B1**

a single straight line would not model the curved trend well, giving poor estimates for the youngest and oldest runners **B1**

Answer: The data does not follow a linear pattern (it curves at the youngest and oldest ages), so a straight line of best fit would not fit the trend well and would give poor estimates at the extremes of age.

Question 6

(a) correct method, e.g. gradient = (90-30)/10 = 6, then 30 + 6 x 12 **M1**

(a) 102 (or over 100) **A1**

(a) explains estimate is unreliable, e.g. 12 hours is outside the range of the data (extrapolation) and/or a test score cannot exceed 100% **B1**

(a) Answer: 102%, which is impossible since a test score cannot exceed 100%; 12 hours is also outside the range of the data collected, so this is extrapolation and unreliable.

(b) correlation does not necessarily mean causation **B1**

(b) gives another possible factor that could affect scores, e.g. natural ability, quality of revision, or prior knowledge **B1**

(b) Answer: The scatter graph only shows correlation, not causation; other factors, such as natural ability or the quality of revision, could also affect test scores.

Question 7

(a) (8, 105) cao **B1**

(a) Answer: (8, 105)

(b) any sensible reason, e.g. the reaction time may have been recorded incorrectly, or this player may have had an unusually fast practice attempt (oe) **B1**

(b) Answer: The reaction time may have been recorded incorrectly, or this player had an unusually fast response on that attempt.

Question 8

(a) 7 cao **B1**

(a) Answer: 7 pages

(b) 36 (minutes) cao **B1**

(b) Answer: 36 minutes

(c) (7, 36) ft from parts (a) and (b) **B1**

(c) Answer: (7, 36)

Question 9

(a) Graph A **B1**

(a) Answer: Graph A

(b) the points in Graph A lie closer to a straight line / are less scattered than the points in Graph B (oe) **B1**

(b) Answer: The points in Graph A lie closer to a straight line than the scattered points in Graph B.

© Revision Library | revisionlibrary.co.uk

Question 10

(a) False **B1**

(a) Answer: False

(b) True **B1**

(b) Answer: True

(c) True **B1**

(c) Answer: True

(d) False **B1**

(d) Answer: False

Question 11

(a) point (26, 44) plotted correctly **B1**

(a) point (28, 48) plotted correctly **B1**

(a) Answer: (26, 44) and (28, 48)

(b) a single straight line, drawn by eye, that follows the trend of all 9 points with a roughly equal number of points on each side **B1**

(b) Answer: A straight line following the trend of the 9 points, e.g. close to the line y = 2x - 8.

(c) 38 (bottles), accept 36 to 40, ft from candidate's line **B1**

(c) Answer: 38 bottles (accept 36 to 40)

Question 12

(a) (3, 8) cao **B1**

(a) Answer: (3, 8)

(b) the outlier does not follow the trend/pattern shown by the rest of the data **B1**

(b) removing it means the line is not pulled away from the trend by this single unusual result, so it better represents the general relationship for the remaining pupils (oe) **B1**

(b) Answer: The outlier does not fit the pattern of the rest of the data, so removing it stops the line of best fit being distorted, giving a line that better represents the true relationship for the other 7 pupils.

(c) 20 hours is far outside the range of data collected (1 to 8 hours), so this would be extrapolation and the trend may not continue (oe) **B1**

(c) Answer: 20 hours is well outside the data collected (1 to 8 hours), so predicting from the line would be extrapolation and may not be reliable.

Question 13

(a) cannot determine whether the estimate is reliable without knowing the strength of the correlation (oe) **B1**

(a) 195 mm is within the observed range of 165 mm to 210 mm, so this is interpolation; it would be more reliable if the points lie close to the line of best fit (oe) **B1**

(a) Answer: 195 mm is within the observed range of 165 mm to 210 mm, so this is interpolation. Its reliability cannot be determined without knowing how closely the points follow the line of best fit.

(b) not reliable / no **B1**

(b) 260 mm is well outside the range of data collected (165 mm to 210 mm), so this is extrapolation and the relationship may not continue in the same way beyond the observed data (oe) **B1**

(b) Answer: Not reliable, because 260 mm is well beyond the range of wing lengths collected (165 mm to 210 mm), so this is extrapolation and the trend may not continue.

© Revision Library | revisionlibrary.co.uk