HIGHER

4.20H Scatter Graphs: Higher Tier Practice

Total Marks

EDEXCEL Edexcel 1MA1 · Calculator allowed · about 65 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 The following pairs of values were plotted on a scatter graph: (1, 2), (2, 3), (3, 5), (4, 6).

a) State whether the correlation is positive, negative or no correlation.

b) Use the two end points (1, 2) and (4, 6) to estimate the gradient of a line of best fit. Show your working.

c) Using the line you estimated in part (b), predict the value of y when x = 6. Give your answer to 2 decimal places.

(a) State whether the correlation is positive, negative or no correlation. **(1)**

(b) Use the two end points (1, 2) and (4, 6) to estimate the gradient of a line of best fit. Show your working. **(2)**

.....

(c) Using the line you estimated in part (b), predict the value of y when x = 6. Give your answer to 2 decimal places. **(1)**

.....

(Total for Question 1 is 4 marks)

2 A scatter graph shows five points: (1, 2), (2, 4), (3, 5), (4, 4), (5, 6).

a) Describe the correlation shown by this data, and explain what this means about the relationship between x and y.

b) A line of best fit is drawn by eye through the points (1, 2) and (5, 6). Find the equation of this line of best fit, in the form $y = mx + c$.

c) Use your line of best fit to estimate y when $x = 8$, and state whether this estimate is found by interpolation or extrapolation.

(a) Describe the correlation shown by this data, and explain what this means about the relationship between x and y. **(2)**

(b) A line of best fit is drawn by eye through the points (1, 2) and (5, 6). Find the equation of this line of best fit, in the form $y = mx + c$. **(2)**

.....

(c) Use your line of best fit to estimate y when $x = 8$, and state whether this estimate is found by interpolation or extrapolation. **(2)**

(Total for Question 2 is 6 marks)

3 The scatter graph shows six points: (1, 2), (2, 4), (3, 5), (4, 4), (5, 6) and (2, 15). A line of best fit for the first five points (ignoring (2, 15)) has equation $y = x + 1$, as found in Question 2.

a) Explain why (2, 15) can be considered an outlier, using the equation of the line of best fit to support your answer.

b) State whether the outlier should normally be included when drawing the line of best fit, and give a reason.

c) A different set of data has five points including (1, 2) and (6, 8), and excludes the outlier. A line of best fit is drawn through (1, 2) and (6, 8). Find the equation of this new line of best fit.

(a) Explain why (2, 15) can be considered an outlier, using the equation of the line of best fit to support your answer. **(2)**

(b) State whether the outlier should normally be included when drawing the line of best fit, and give a reason. **(1)**

(c) A different set of data has five points including (1, 2) and (6, 8), and excludes the outlier. A line of best fit is drawn through (1, 2) and (6, 8). Find the equation of this new line of best fit. **(3)**

.....

(Total for Question 3 is 6 marks)

4 A scatter graph plots hours of sunshine (x) against ice lolly sales in tens (y), giving the points (1, 3), (2, 5), (3, 6), (4, 9).

a) A line of best fit is drawn by eye through the points (1, 3) and (4, 9). Find the equation of this line, in the form $y = mx + c$.

b) Use your line of best fit to estimate the sales when there are 3.5 hours of sunshine.

c) Use your line of best fit to estimate the number of hours of sunshine when sales are 15 (tens).

d) Explain whether your answer to part (c) is likely to be reliable.

(a) A line of best fit is drawn by eye through the points (1, 3) and (4, 9). Find the equation of this line, in the form $y = mx + c$. **(3)**

.....

(b) Use your line of best fit to estimate the sales when there are 3.5 hours of sunshine. **(2)**

(c) Use your line of best fit to estimate the number of hours of sunshine when sales are 15 (tens). **(2)**

(d) Explain whether your answer to part (c) is likely to be reliable. **(1)**

(Total for Question 4 is 8 marks)

5 A researcher plots the number of ice creams sold (x) against the number of drownings recorded at a beach (y) over several summer days, where x ranges from 5 to 20. She finds a strong positive correlation, with a line of best fit $y = 0.4x + 1$.

a) State what a strong positive correlation between x and y means.

b) The researcher concludes that selling more ice creams causes more drownings. Explain why this conclusion is not necessarily valid, suggesting a possible reason for the association.

c) Explain whether the line of best fit is likely to give a reliable estimate for the number of drownings when $x = 3$, given that the original data only covers $x = 5$ to $x = 20$.

d) A second scatter graph, plotting the same drowning data against daily temperature, shows an even stronger positive correlation than with ice cream sales. Suggest, with a reason, why temperature might be a more sensible variable to investigate as a genuine underlying cause.

(a) State what a strong positive correlation between x and y means. **(2)**

(b) The researcher concludes that selling more ice creams causes more drownings. Explain why this conclusion is not necessarily valid, suggesting a possible reason for the association. **(3)**

(c) Explain whether the line of best fit is likely to give a reliable estimate for the number of drownings when $x = 3$, given that the original data only covers $x = 5$ to $x = 20$. **(2)**

(d) A second scatter graph, plotting the same drowning data against daily temperature, shows an even stronger positive correlation than with ice cream sales. Suggest, with a reason, why temperature might be a more sensible variable to investigate as a genuine underlying cause. **(1)**

(Total for Question 5 is 8 marks)

6 A spring is stretched by hanging different loads on it. The load x (newtons) and the total length y (cm) of the spring were recorded for loads between 2 N and 8 N: (2, 5), (4, 9), (6, 13), (8, 17).

a) Show that these four points lie exactly on a straight line, and find its equation in the form $y = mx + c$.

b) Interpret the gradient and the y-intercept of this line in the context of the spring.

c) The manufacturer states the spring will break if stretched beyond 25 cm. Use the equation to find the load at which this would happen, and explain whether your answer involves interpolation or extrapolation.

d) A second, stiffer spring gives the data points (1, 3), (3, 6), (5, 9) for loads between 1 N and 5 N. Show that this data also lies exactly on a straight line, find its equation, and hence find the load at which both springs would have the same total length.

(a) Show that the four points (2, 5), (4, 9), (6, 13), (8, 17) lie exactly on a straight line, and find its equation in the form $y = mx + c$. **(4)**

(b) Interpret the gradient and the y-intercept of this line in the context of the spring. **(2)**

(c) The manufacturer states the spring will break if stretched beyond 25 cm. Use the equation to find the load at which this would happen, and explain whether your answer involves interpolation or extrapolation. **(4)**

(d) A second, stiffer spring gives the data points (1, 3), (3, 6), (5, 9) for loads between 1 N and 5 N. Show that this data also lies exactly on a straight line, find its equation, and hence find the load at which both springs would have the same total length. **(6)**

.....

(Total for Question 6 is 16 marks)

Mark scheme · 4.20H Scatter Graphs: Higher Tier Practice

Question 1

- (a) **B1** positive correlation cao
- **(a) Answer: Positive correlation**
- (b) **M1** method: gradient = (change in y)/(change in x) using two given points
- (b) **A1** 4/3 or 1.333... cao
- **(b) Answer: 4/3 or 1.333...**
- (c) **B1** 8.67 cao (answer to 2 dp)
- **(c) Answer: 8.67**

Question 2

- (a) **M1** identify positive correlation
- (a) **A1** explain in context: as x increases, y tends to increase oe
- **(a) Answer: Positive correlation: as x increases, y tends to increase.**
- (b) **M1** gradient = (6-2)/(5-1) = 1
- (b) **A1** y = x + 1 cao
- **(b) Answer: y = x + 1**
- (c) **M1** substitute x = 8 into y = x + 1 to get y = 9
- (c) **A1** state extrapolation, since x = 8 is outside the original data range (1 to 5), so the estimate may be less reliable cao
- **(c) Answer: y = 9; this is extrapolation, since x = 8 lies outside the given data range of 1 to 5, so the estimate may not be reliable.**

Question 3

- (a) **M1** the line of best fit predicts y = 2 + 1 = 3 when x = 2
- (a) **A1** the actual value (15) is far from the predicted value (3), much further than any other point, so (2, 15) is an outlier cao
- **(a) Answer: The line predicts y = 3 at x = 2, but the actual value is 15, a much larger difference than for any other point, so (2, 15) is an outlier.**
- (b) **B1** no, outliers are usually excluded (or investigated separately) since including them would distort the line of best fit away from the general trend of the other points oe
- **(b) Answer: No; an outlier is usually excluded (or investigated separately), since including it would pull the line of best fit away from the general trend shown by the rest of the data.**
- (c) **M1** gradient = (8-2)/(6-1) = 6/5 = 1.2
- (c) **M1** use point (1,2) to find intercept: 2 = 1.2*1 + c
- (c) **A1** y = 1.2x + 0.8 cao
- **(c) Answer: y = 1.2x + 0.8**

Question 4

- (a) **M1** gradient = (9-3)/(4-1) = 2
- (a) **M1** use point (1,3) to find intercept: 3 = 2*1 + c
- (a) **A1** y = 2x + 1 cao
- **(a) Answer: y = 2x + 1**
- (b) **M1** substitute x = 3.5 into y = 2x + 1
- (b) **A1** 8 (tens of ice lollies) cao
- **(b) Answer: 8 (i.e. 80 ice lollies)**
- (c) **M1** set 15 = 2x + 1 and rearrange
- (c) **A1** x = 7 cao
- **(c) Answer: 7 hours**
- (d) **B1** not necessarily reliable; x = 7 is outside the original data range (1 to 4 hours), so this is extrapolation and the pattern may not continue oe
- **(d) Answer: Not necessarily reliable, since x = 7 is outside the original data range (1 to 4 hours) - this is extrapolation, and the relationship may not continue in the same way beyond the observed data.**

Question 5

- (a) **M1** as x increases, y also tends to increase
- (a) **A1** the points lie close to a straight line (little scatter), so the association is strong oe
- **(a) Answer: As x increases, y tends to increase, and the points lie close to a straight line, showing a strong association.**
- (b) **M1** state that correlation does not imply causation
- (b) **A1** identify that another factor could be responsible for both variables (a confounding variable)
- (b) **A1** give a plausible reason: hot weather leads to more ice cream sales and, separately, more people swimming (so more risk of drowning), rather than ice cream sales directly causing drownings oe
- **(b) Answer: Correlation does not imply causation. A third factor, such as hot weather, could cause both more ice cream sales and more people swimming (and so more drownings), without ice cream sales directly causing drownings.**
- (c) **M1** recognise that x = 3 lies outside the original data range (5 to 20), so this would be extrapolation
- (c) **A1** state that extrapolated estimates are less reliable, since the linear relationship may not hold outside the range of the observed data oe
- **(c) Answer: Not reliable: x = 3 is outside the data range (5 to 20), so using the line here is extrapolation, and the relationship may not hold outside the observed range.**
- (d) **B1** temperature can plausibly cause both more ice cream sales and more swimming directly, so it is a more sensible candidate for a genuine underlying cause than ice cream sales themselves oe
- **(d) Answer: Temperature is more plausible as a genuine cause, since hot weather can directly explain both more ice cream sales and more people going swimming, whereas ice cream sales have no direct mechanism for causing drownings.**

Question 6

- (a) **M1** gradient using (2,5) and (4,9): (9-5)/(4-2) = 2
- (a) **A1** use (2,5) to find intercept: 5 = 2*2 + c, so c = 1, giving y = 2x + 1
- (a) **M1** check (6,13): 2*6 + 1 = 13, which matches
- (a) **A1** check (8,17): 2*8 + 1 = 17, which matches, so all four points lie exactly on y = 2x + 1 cao
- **(a) Answer: y = 2x + 1; all four points lie exactly on this line.**
- (b) **M1** gradient = 2 means the spring stretches by an extra 2 cm for every extra 1 N of load applied
- (b) **A1** y-intercept = 1 means the spring's natural (unloaded) length is 1 cm, when x = 0 oe
- **(b) Answer: Gradient 2: the length increases by 2 cm for each extra newton of load. Intercept 1: the spring's natural length (with no load) is 1 cm.**
- (c) **M1** set 25 = 2x + 1 and rearrange
- (c) **A1** x = 12 N cao
- (c) **M1** state the original data only covers loads from 2 N to 8 N
- (c) **A1** conclude x = 12 N is outside this range, so this is extrapolation and the prediction may be unreliable (e.g. the spring may not obey the same linear relationship close to breaking point) oe
- **(c) Answer: x = 12 N. This is extrapolation, since 12 N is outside the original data range of 2 N to 8 N, so the prediction may not be reliable.**
- (d) **M1** gradient using (1,3) and (3,6): (6-3)/(3-1) = 1.5
- (d) **A1** use (1,3) to find intercept: 3 = 1.5*1 + c, so c = 1.5, giving y = 1.5x + 1.5
- (d) **M1** check (5,9): 1.5*5 + 1.5 = 9, which matches, so all three points lie exactly on this line
- (d) **M1** set the two spring equations equal: 2x + 1 = 1.5x + 1.5
- (d) **A1** x = 1.0 N cao
- (d) **A1** corresponding length y = 2(1) + 1 = 3 cm, so both springs are 3 cm long when the load is 1.0 N cao
- **(d) Answer: Second spring: y = 1.5x + 1.5. Both springs have the same length (3 cm) when the load is 1.0 N.**