



S31-30min Scatter Graphs and Line of Best Fit: 30-minute Pack

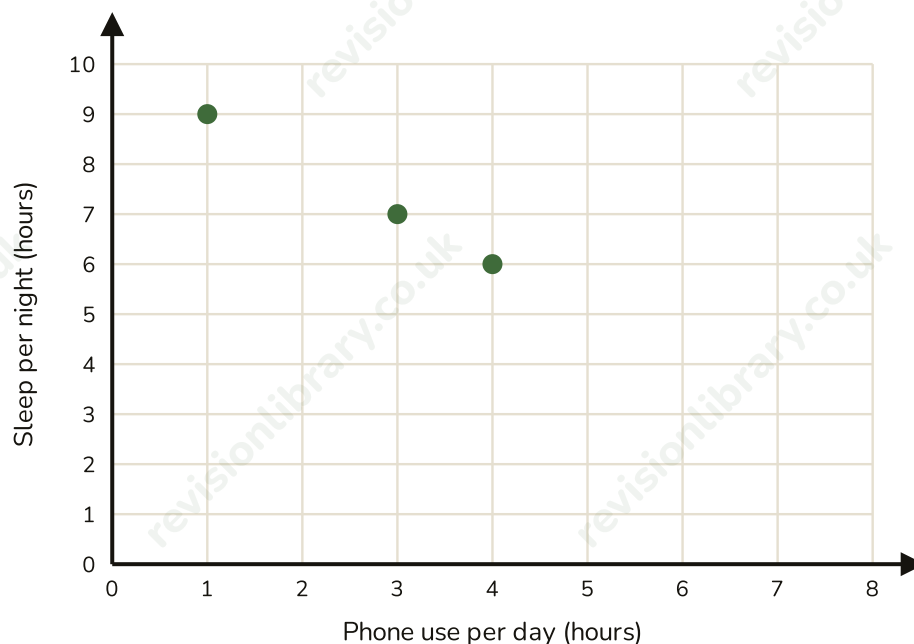
EDEXCEL 1ST0 • Calculator allowed • about 30 minutes

Total Marks

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** A student wants to investigate whether there is a relationship between the time Year 11 students spend on phone apps each day and the amount of sleep they get each night.



- (a) Plan: Write down a suitable hypothesis for this investigation. (1)
-
- (b) Collect: The school has 220 Year 11 students split across 8 tutor groups of different sizes. Name a sampling method the student could use to choose 20 students for the survey, and give a reason for your choice. (2)

- (c) Process: The table shows the results for two more students.

(2)

Student	D	E
Phone use (hours)	6	7
Sleep (hours)	5	4

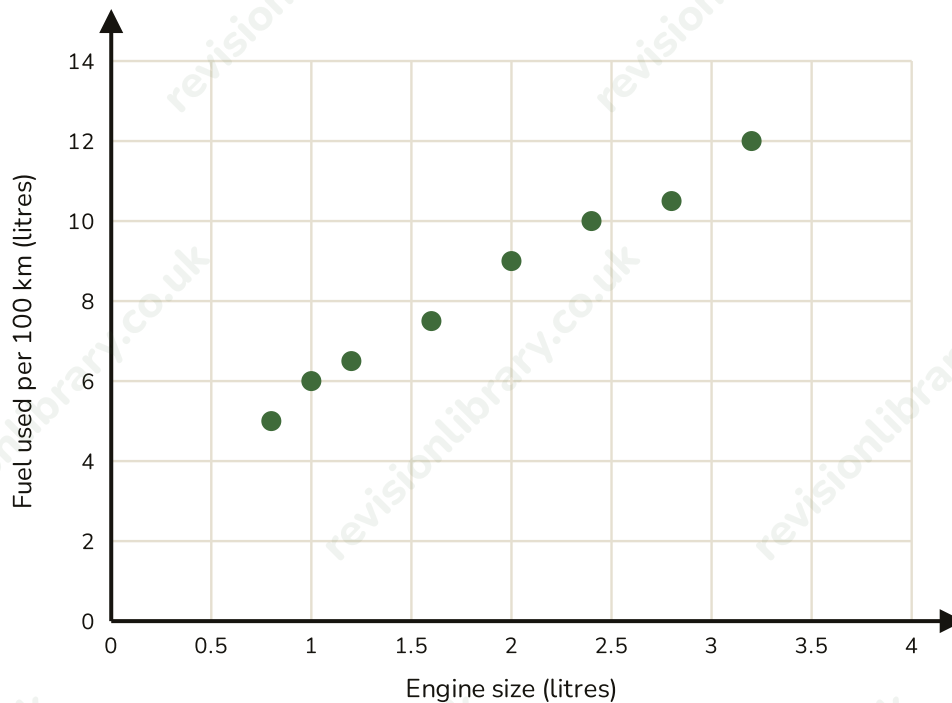
Plot the points for students D and E on the grid.

- (d) Interpret: The completed scatter graph shows a negative correlation. Write a conclusion for the investigation, including a comment on causation.

(2)

(Total for Question 1 is 7 marks)

- 2 The scatter graph shows, for 8 cars, the engine size and the fuel used per 100 km driven. The points are plotted but no line of best fit has been drawn.



- (a) Write down the type of correlation shown by the scatter graph. (1)
- (b) On the grid, draw a line of best fit for this data. (2)
- (c) Use your line of best fit to estimate the fuel used per 100 km for a car with an engine size of 1.8 litres. (2)
- (d) Explain why it would not be appropriate to use your line of best fit to predict the fuel used per 100 km for a car with a 6.0 litre engine. (2)

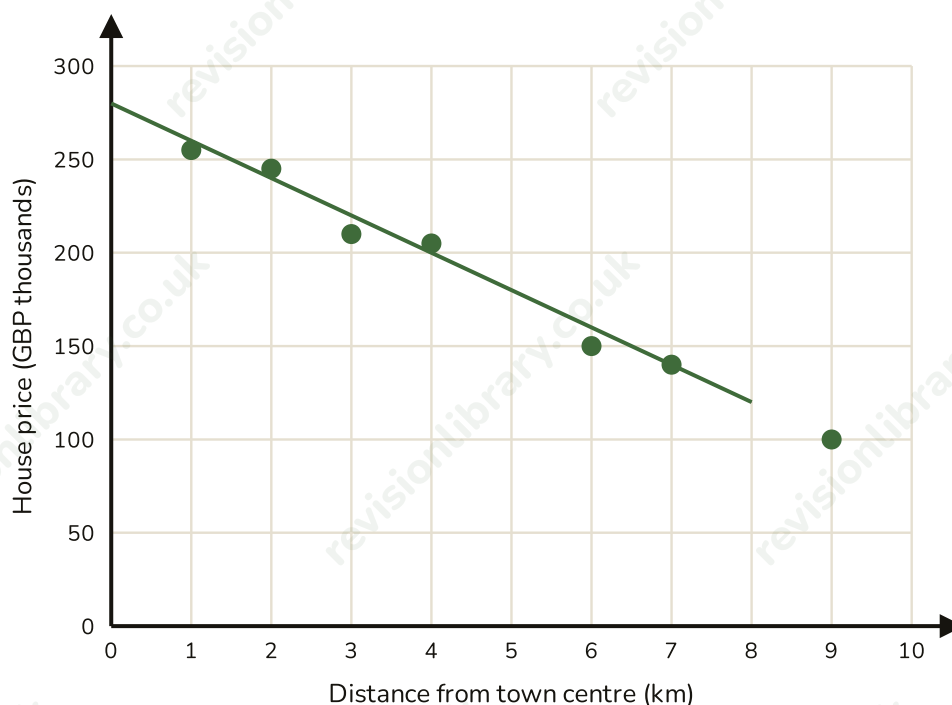
(Total for Question 2 is 7 marks)

3 A gardener records the height of a sunflower plant every week for 20 weeks. The equation of the line of best fit for the data is $h = 1.5w + 8$, where w is the number of weeks since planting and h is the height of the plant in cm.

- (a) Use the equation to estimate the height of the plant after 14 weeks. (2)
- (b) Interpret what the gradient (1.5) and the intercept (8) tell you about the plant's growth. (2)
- (c) Explain why it would not be sensible to use this equation to estimate the height of the plant after 100 weeks. (1)

(Total for Question 3 is 5 marks)

- 4 The scatter graph shows, for 7 houses, the distance from the town centre and the house price, together with a line of best fit that passes through the points (0, 280) and (8, 120).



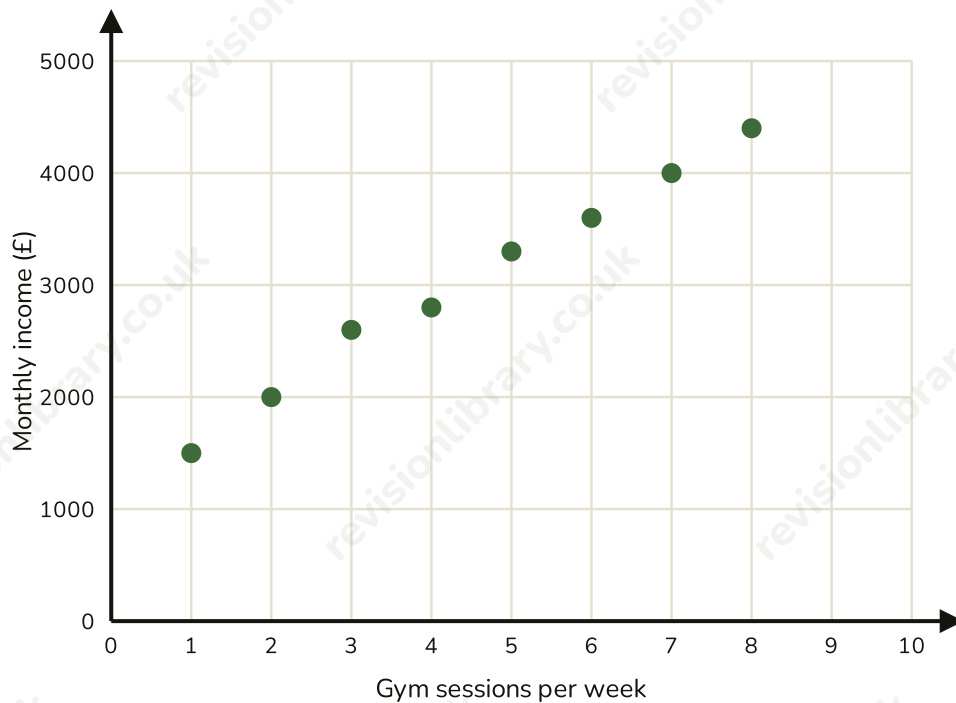
- (a) Find the gradient of the line of best fit, using the points (0, 280) and (8, 120). (2)

- (b) Hence write down the equation of the line of best fit in the form $p = md + c$, where d is the distance from the town centre and p is the house price in GBP thousands. (1)

- (c) Use the equation to estimate the house price for a home 5 km from the town centre. (1)

(Total for Question 4 is 4 marks)

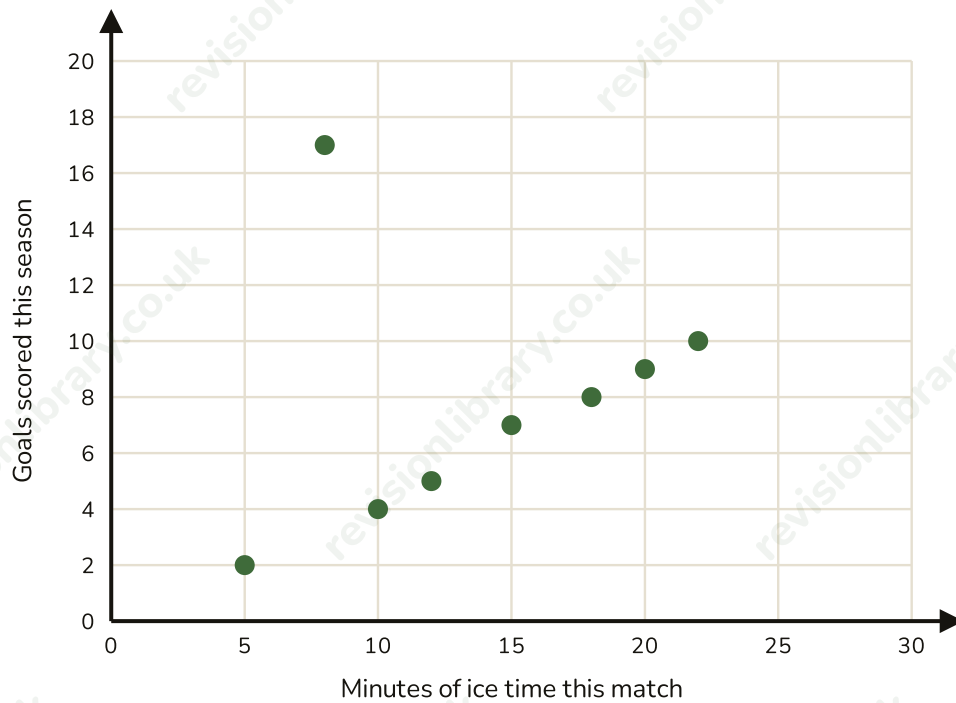
- 5 A personal trainer records, for 8 gym members, the number of gym sessions attended per week and their monthly income.



- (a) Describe the correlation shown by the scatter graph. (1)
- (b) The trainer claims: 'This proves that going to the gym more often causes people to earn more money.' Explain why this conclusion is not necessarily correct, and suggest one other factor that could explain the pattern. (2)
- (c) The trainer extrapolates the line of best fit to claim that someone attending 20 sessions a week would earn about £11,000 a month. Explain why this estimate is unreliable. (2)

(Total for Question 5 is 5 marks)

- 6** The scatter graph shows, for 8 players in an ice hockey team, the number of minutes of ice time in a match and the number of goals scored so far this season.



- (a) Identify the coordinates of the outlier, and suggest a possible reason for it. (2)
- (b) Explain whether it would be sensible to exclude this point before drawing a line of best fit for the rest of the team. (1)
- (c) Using a line of best fit drawn for the remaining 7 (non-outlier) points, estimate the number of goals scored by a player with 16 minutes of ice time. (2)

(Total for Question 6 is 5 marks)

Mark scheme · S31-30min Scatter Graphs and Line of Best Fit: 30-minute Pack

Question 1

- (a) **B1** a testable statement linking the two variables, e.g. students who spend more time on phone apps tend to get less sleep, oe
- (a) **Answer: Students who spend more time on phone apps each day tend to get less sleep at night.**
- (b) **B1** Stratified sampling (by tutor group), oe
- (b) **B1** valid reason, e.g. ensures each tutor group is represented in proportion to its size, giving a more representative sample, oe
- (b) **Answer: Stratified sampling by tutor group, because it ensures each group is represented in proportion to its size, giving a more representative sample.**
- (c) **B1** point for D plotted correctly at (6,5) ($\pm$ half a small square)
- (c) **B1** point for E plotted correctly at (7,4) ($\pm$ half a small square)
- (c) **Answer: Points plotted at (6,5) for D and (7,4) for E.**
- (d) **B1** correct context conclusion, e.g. students who spend more time on phone apps tend to get less sleep, oe
- (d) **B1** caution about causation, e.g. this does not prove phone use causes less sleep - another factor, such as staying up late generally, could affect both, oe
- (d) **Answer: Students who spend longer on phone apps tend to sleep less, but this correlation does not prove phone use causes less sleep - another factor, such as generally staying up later, could affect both.**

Question 2

- (a) **B1** Positive correlation cao
- (a) **Answer: Positive correlation**
- (b) **B1** a single ruled straight line drawn across the full range of the plotted points
- (b) **B1** line follows the trend of the data with a roughly equal number of points on each side, e.g. close to a line from about (0.8,5) to (3.2,12)
- (b) **Answer: A ruled line running from approximately (0.8,5) to (3.2,12).**
- (c) **M1** reads a value from a single ruled line of best fit at $x = 1.8$
- (c) **A1** answer consistent with their line; expected value approximately 8 litres (accept 7 to 9), ft
- (c) **Answer: Approximately 8 litres (accept 7-9)**
- (d) **B1** identifies that 6.0 litres is well outside the range of the original data (0.8 to 3.2 litres), so this would be extrapolation
- (d) **B1** correct conclusion, e.g. the relationship may not stay linear for such a large engine, so the estimate would be unreliable, oe
- (d) **Answer: Unreliable - 6.0 litres is far outside the 0.8 to 3.2 litre range of the data, so this is extrapolation and the same linear pattern may not continue.**

Question 3

- (a) **M1** substitutes $w = 14$ into $h = 1.5w + 8$, oe
- (a) **A1** 29 cm cao
- (a) **Answer: 29 cm**
- (b) **B1** gradient: the plant grows by about 1.5 cm each week, oe
- (b) **B1** intercept: the plant's height at the time of planting ($w = 0$) was about 8 cm, oe
- (b) **Answer: The gradient shows the plant grows about 1.5 cm per week; the intercept shows the plant was about 8 cm tall when it was planted.**
- (c) **B1** 100 weeks is far outside the 20-week range of the data (extrapolation), so the plant's growth pattern may no longer be linear (e.g. it may have stopped growing or died), oe
- (c) **Answer: 100 weeks is far beyond the 20-week study, so this would be extrapolation and the plant's growth would very likely no longer follow the same straight-line pattern.**

Question 4

- (a) **M1** gradient = $(120 - 280) / (8 - 0)$, oe
- (a) **A1** -20 cao
- (a) **Answer: -20**
- (b) **B1** $p = -20d + 280$ cao, ft from (a)
- (b) **Answer: $p = -20d + 280$**
- (c) **B1** 180 (GBP thousands) cao, ft from (b)
- (c) **Answer: £180,000**

Question 5

- (a) **B1** positive correlation, oe
- (a) **Answer: Positive correlation**
- (b) **B1** states that correlation does not prove causation, oe
- (b) **B1** valid confounding factor, e.g. people with flexible, well-paid jobs may have more free time and money to spend on both gym sessions and training, oe
- (b) **Answer: Correlation does not prove causation. People with well-paid, flexible jobs may simply have more free time and money to spend on the gym, rather than the gym causing higher income.**
- (c) **B1** identifies that 20 sessions is far outside the range of the data (1 to 8 sessions), so this is extrapolation
- (c) **B1** correct supporting reason, e.g. the pattern may not continue at such extreme values, and attending 20 sessions a week is not realistically possible, oe
- (c) **Answer: Unreliable - 20 sessions a week is far outside the 1 to 8 session range of the data, so this is extrapolation, and the linear pattern (and the claim itself) is unlikely to hold at such an extreme value.**

Question 6

- (a) **B1** (8, 17) cao
- (a) **B1** plausible reason, e.g. this player may be an unusually prolific scorer who makes the most of limited ice time, or benefited from lucky rebounds, oe
- (a) **Answer: (8, 17) - possibly an unusually prolific scorer, or a run of fortunate goals in limited playing time.**
- (b) **B1** yes, with valid reason, e.g. keeping it would distort the line away from the trend followed by the other 7 players, oe
- (b) **Answer: Yes - keeping it in would pull the line of best fit away from the trend shown by the other 7 players.**
- (c) **M1** reads a value from a sensible line of best fit through the 7 remaining points at $x = 16$, ignoring (8,17)
- (c) **A1** answer consistent with their line; expected value approximately 7.5 (accept 7 to 8), ft
- (c) **Answer: Approximately 7-8 goals**