BIVARIATE STATISTICS

A BIVARIATE VARIABLES

A.1 IDENTIFYING EXPLANATORY AND RESPONSE VARIABLES

Ex 1: Scenario: A botanist explores the effect of daily sunlight on plant growth by measuring the average hours of sunlight (x) and the resulting height increase in cm (y).
For this study, identify the role of each variable:

- "Sunlight Hours per Day" is the:
  ☐ independent variable (x)
  ☐ dependent variable (y)
- "Height Increase (cm)" is the:
  ☐ independent variable (x)
  ☐ dependent variable (y)

Ex 2: Scenario: A fitness coach examines how weekly exercise time impacts a client's resting heart rate (bpm).
For this study, identify the role of each variable:

- "Resting Heart Rate (bpm)" is the:
  ☐ independent variable (x)
  ☐ dependent variable (y)
- "Exercise Hours per Week" is the:
  ☐ dependent variable (y)
  ☐ independent variable (x)

Ex 3: Scenario: A baker studies the impact of oven temperature (°C) on the final height of her bread (cm).
For this study, identify the role of each variable:

- "Oven Temperature (°C)" is the:
  ☐ dependent variable (y)
  ☐ independent variable (x)
- "Bread Height (cm)" is the:
  ☐ dependent variable (y)
  ☐ independent variable (x)

Ex 4: Scenario: A student investigates how study time influences quiz scores. He tracks hours spent studying per week and the scores achieved (out of 20).
For this study, identify the role of each variable:

- "Quiz Score (out of 20)" is the:
  ☐ dependent variable (y)
  ☐ independent variable (x)
- "Study Hours per Week" is the:
  ☐ independent variable (x)
  ☐ dependent variable (y)

B SCATTER PLOTS

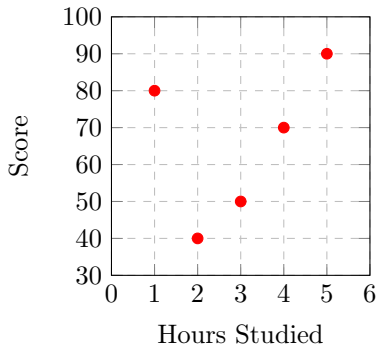
B.1 MATCHING DATA TABLES TO SCATTER PLOTS

MCQ 5: The table below shows hours studied (x) and exam scores (y) for five students.

Student	O	P	Q	R	S
x: Hours Studied	1	2	4	3	5
y: Score Obtained	40	50	80	70	90

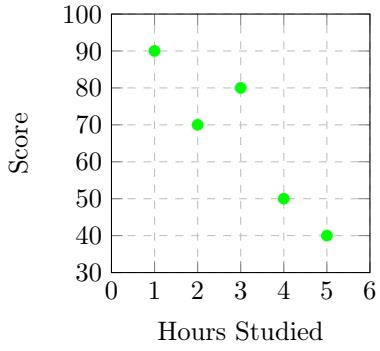
Which scatter plot correctly represents the data in the table?

Graph E



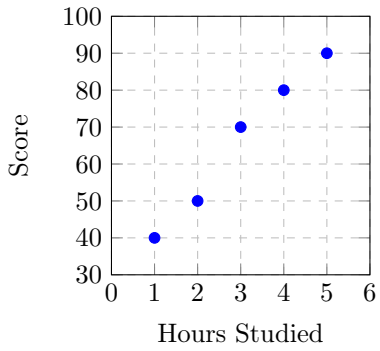
☐

Graph F



☐

Graph G



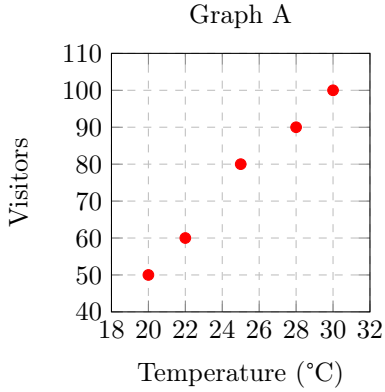
☐

MCQ 6: The table below shows daily temperature (x, in °C) and beach visitors (y).

x: Temperature (°C)	20	25	22	28	30
y: Beach Visitors	50	80	60	90	100

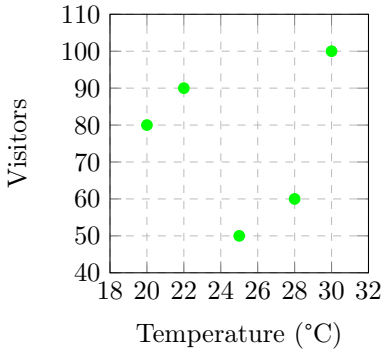
Which scatter plot correctly represents the data?

Graph A



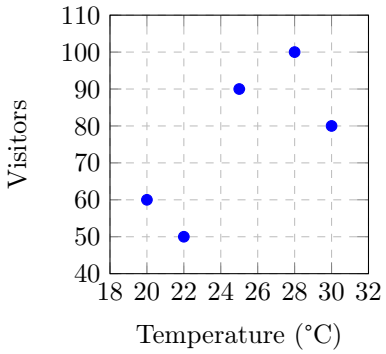
☐

Graph B



□

Graph C



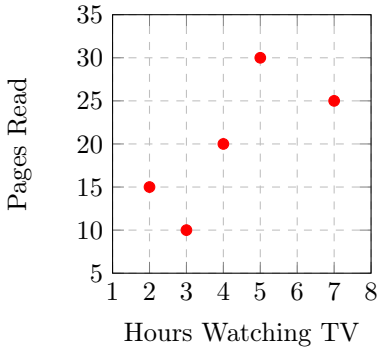
□

MCQ 7: The table shows hours spent watching TV (x) and pages read (y).

x: Hours Watching TV	2	5	3	7	4
y: Pages Read	30	15	25	10	20

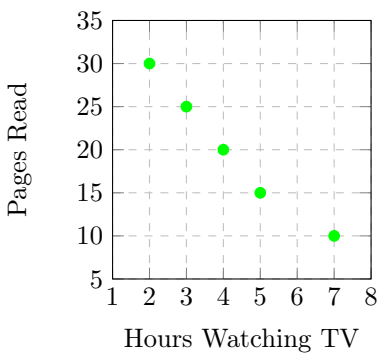
Which scatter plot correctly represents the data?

Graph P



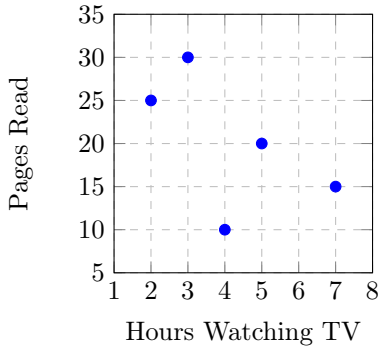
□

Graph Q



□

Graph R



□

B.2 PLOTTING SCATTER PLOTS

Ex 8: Scenario: A botanist measures the age and height of several trees of the same species. The collected data is shown in the table below.

Age (years) (x)	2	5	8	10	12	15
Height (m) (y)	3	6	9	11	14	16

Construct a scatter plot to represent this data. Ensure you label the axes and choose an appropriate scale.

Ex 9: Scenario: A consumer group rates several brands of chocolate based on taste (out of 10) and records their price. The data is shown in the table below.

Price ($) ($x$)	1	2	3	4	5
Taste Rating (/10) (y)	4	7	6	8	10

Construct a scatter plot to represent this data. Let Price be the independent variable. Ensure you label the axes and choose an appropriate scale.

Ex 10: Scenario: A health researcher investigates a possible link between daily coffee consumption and hours of sleep. The data for six individuals is shown below.

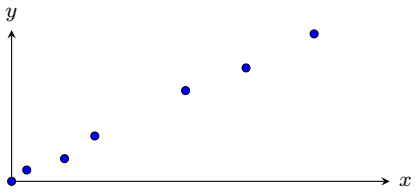
Coffee (cups per day) (x)	0	1	2	3	4	5
Sleep (hours per night) (y)	7.5	8.5	8.0	6.0	7.0	6.5

Construct a scatter plot to represent this data. Let coffee consumption be the independent variable. Ensure you label the axes and choose an appropriate scale.

C CORRELATION

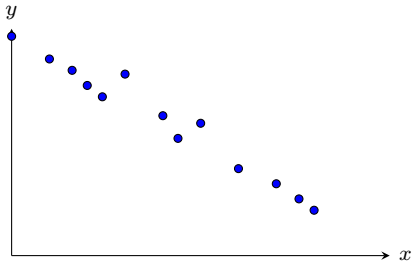
C.1 IDENTIFYING THE DIRECTION OF A CORRELATION

Ex 11: Determine the direction of the correlation shown in the scatter plot below.



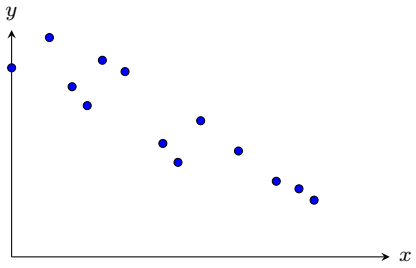
The correlation is ☐ positive ☐ negative .

Ex 12: Determine the direction of the correlation shown in the scatter plot below.



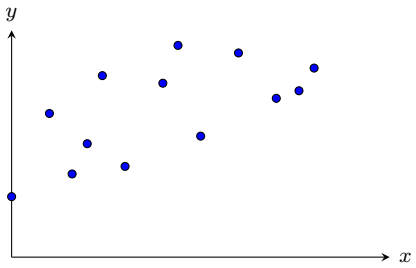
The correlation is ☐ positive ☐ negative .

Ex 13: Determine the direction of the correlation shown in the scatter plot below.



The correlation is ☐ positive ☐ negative .

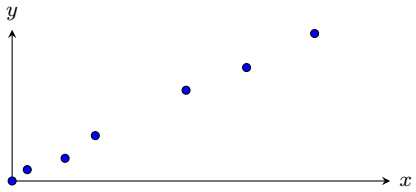
Ex 14: Determine the direction of the correlation shown in the scatter plot below.



The correlation is ☐ positive ☐ negative .

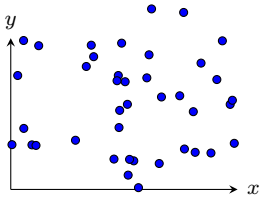
C.2 IDENTIFYING THE STRENGTH OF A CORRELATION

Ex 15: Determine the strength of the correlation shown in the scatter plot below.



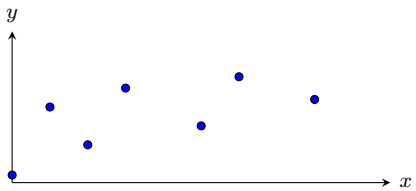
The correlation strength is ☐ Strong ☐ Weak ☐ None .

Ex 16: Determine the strength of the correlation shown in the scatter plot below.



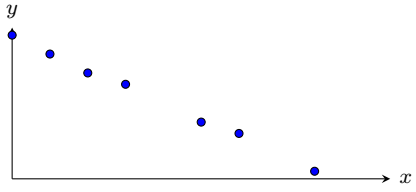
The correlation strength is ☐ Strong ☐ Weak ☐ None .

Ex 17: Determine the strength of the correlation shown in the scatter plot below.



The correlation strength is ☐ Strong ☐ Weak ☐ None .

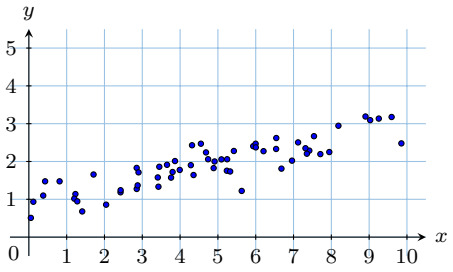
Ex 18: Determine the strength of the correlation shown in the scatter plot below.



The correlation strength is ☐ Strong ☐ Weak ☐ None .

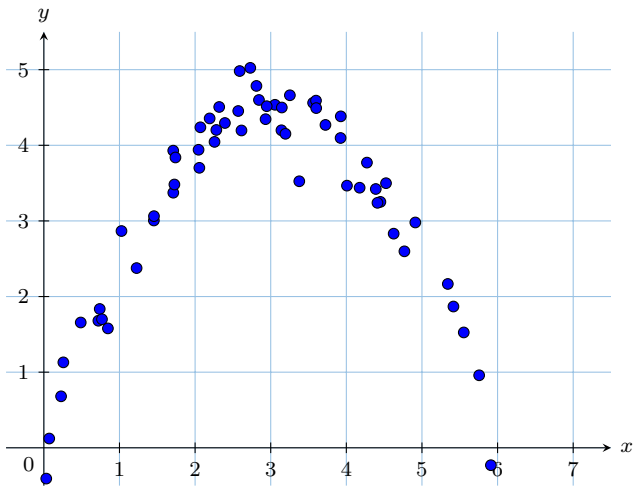
C.3 IDENTIFYING THE FORM OF A CORRELATION

Ex 19: Determine if the form of the correlation is linear or non-linear.



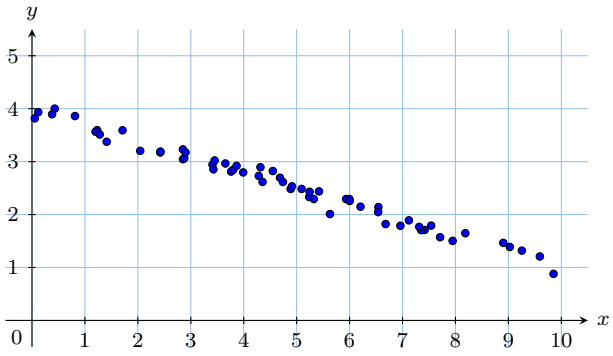
The correlation is ☐ linear ☐ non-linear .

Ex 20: Determine if the form of the correlation is linear or non-linear.



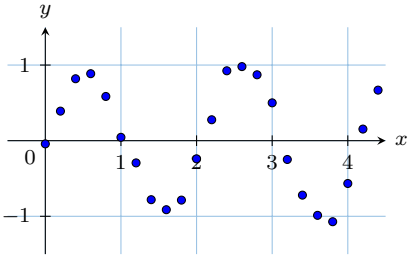
The correlation is ☐ linear ☐ non-linear .

Ex 21: Determine if the form of the correlation is linear or non-linear.



The correlation is ☐ linear ☐ non-linear .

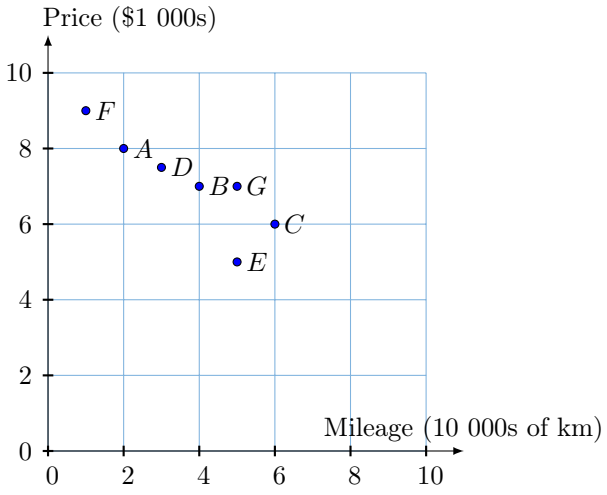
Ex 22: Determine if the form of the correlation is linear or non-linear.



The correlation is ☐ linear ☐ non-linear .

C.4 INTERPRETING OUTLIERS IN CONTEXT

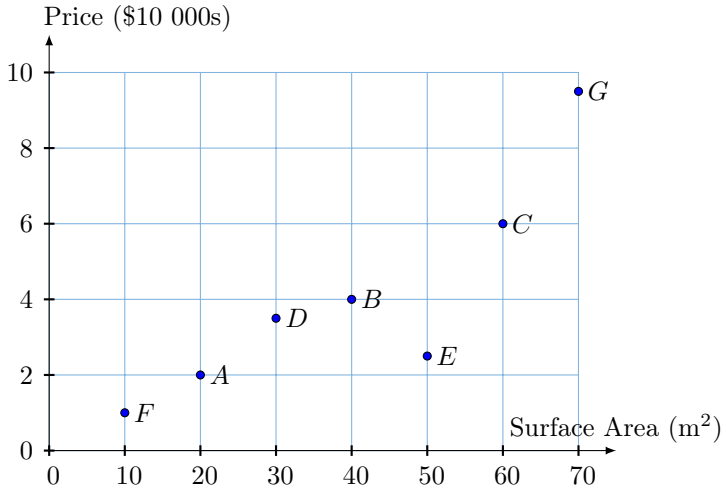
Ex 23: Scenario: You are analyzing data on used cars, plotting price versus mileage.



An outlier that represents a "good deal" would be a car that is underpriced for its mileage. Identify this car.

Select one: ☐ A ☐ B ☐ E ☐ G ☐ D .

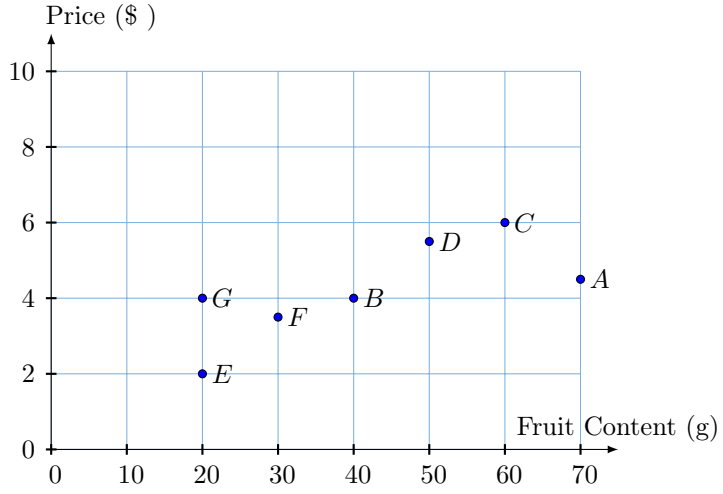
Ex 24: Scenario: You are analyzing house prices versus their surface area (in square meters).



Which house is an outlier that represents a "good deal" (i.e., is underpriced for its size)?

Select one: ☐ A ☐ B ☐ E ☐ G ☐ D .

Ex 25: Scenario: You are researching jam brands, plotting price versus fruit content (in grams).

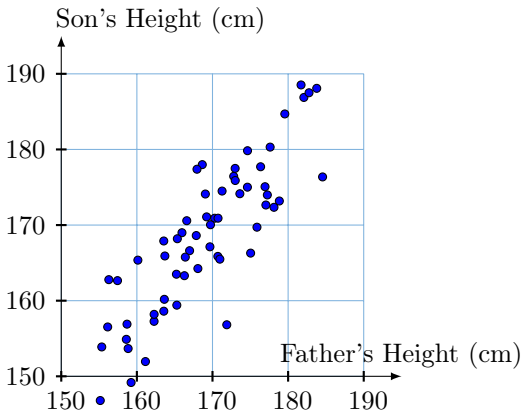


Which jam is an outlier that represents a "good deal" (a low price for a high fruit content)?

Select one: ☐ A ☐ B ☐ E ☐ G ☐ D .

C.5 INTERPRETING TRENDS IN SCATTER PLOTS

MCQ 26: Scenario: A researcher explores if sons of taller fathers tend to be taller. The scatter plot shows the relationship between fathers' heights and their sons' heights.



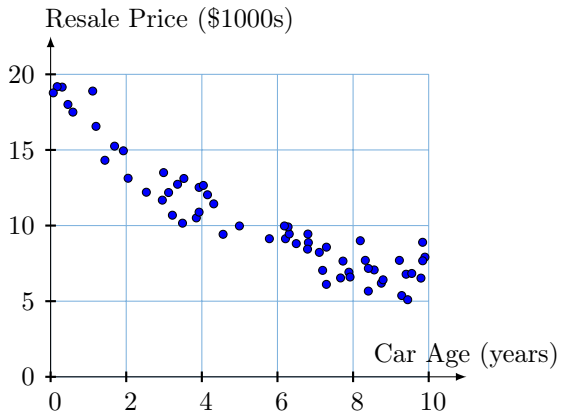
Which statement best describes the association shown in the plot?

☐ Taller fathers tend to have taller sons.

☐ Taller fathers tend to have shorter sons.

☐ There is no clear relationship between fathers' and sons' heights.

MCQ 27: Scenario: A study analyzes how the age of a used car affects its resale price.



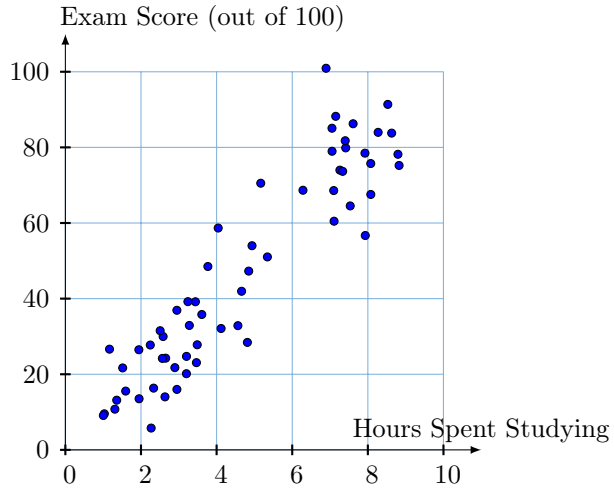
Which statement best describes the association shown in the plot?

☐ Older cars tend to have higher resale prices.

☐ Older cars tend to have lower resale prices.

☐ There is no clear relationship between car age and resale price.

MCQ 28: **Scenario:** An investigation explores the relationship between hours spent studying and exam scores.



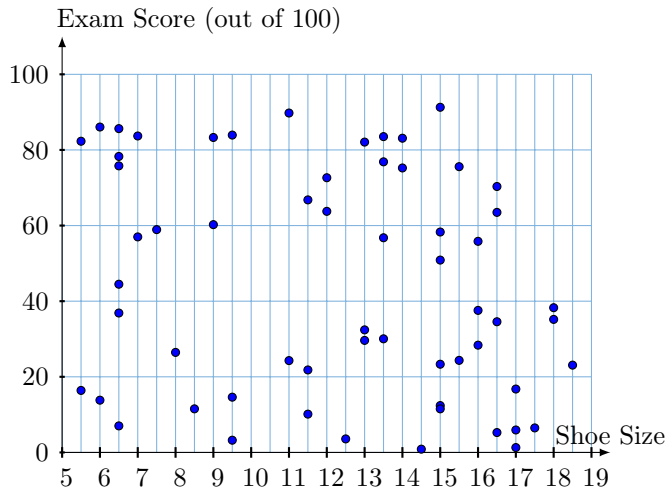
Which statement best describes the association shown in the plot?

☐ Students who study more hours tend to score higher on the exam.

☐ Students who study more hours tend to score lower on the exam.

☐ There is no clear relationship between study hours and exam scores.

MCQ 29: **Scenario:** A researcher investigates whether a student's shoe size relates to their exam score.



Which statement best describes the association shown in the plot?

☐ Students with larger shoe sizes tend to score higher on the exam.

☐ Students with larger shoe sizes tend to score lower on the exam.

☐ There is no clear relationship between shoe size and exam score.

D CORRELATION VS. CAUSATION

D.1 DISTINGUISHING BETWEEN CORRELATION AND CAUSATION

Ex 30: **Scenario:** Consider the relationships between the following lifestyle factors.
For each pair, determine if the relationship is most likely a direct causation or a correlation.

- Eating a nutritious diet and having energy: ☐ Correlation ☐ Causation
- Having energy and living a long life: ☐ Correlation ☐ Causation

Ex 31: **Scenario:** Consider the events that occur on a hot summer day.
For each pair, determine if the relationship is a causation or a correlation.

- Sunny weather and eating ice cream: ☐ Correlation ☐ Causation
- Eating ice cream and getting sunburned: ☐ Correlation ☐ Causation
- Sunny weather and getting sunburned: ☐ Correlation ☐ Causation

Ex 32: **Scenario:** Consider the habits and characteristics of students.
For each pair, determine if the relationship is most likely a causation or a correlation.

- Studying hard and getting good grades: ☐ Correlation ☐ Causation
- Getting good grades and wearing glasses: ☐ Correlation ☐ Causation

D.2 DISTINGUISHING BETWEEN CORRELATION AND CAUSATION

Ex 33: A study found a strong positive correlation ($r = 0.85$) between the number of firefighters present at a fire scene (x) and the amount of damage caused by the fire in dollars (y).

1. Explain whether this correlation implies that having more firefighters *causes* more damage.

2. Identify a likely confounding variable that influences both the number of firefighters deployed and the amount of damage caused.

Ex 34: A statistical study of elementary school students shows a strong positive correlation between shoe size (x) and reading comprehension scores (y).

1. Explain whether this correlation implies that having larger feet *causes* an improvement in reading skills.

2. Identify a likely confounding variable that explains why these two variables increase together.

Ex 35: Data collected from a coastal city over several years reveals a strong positive correlation ($r = 0.9$) between monthly ice cream sales (x) and the number of shark attacks (y).

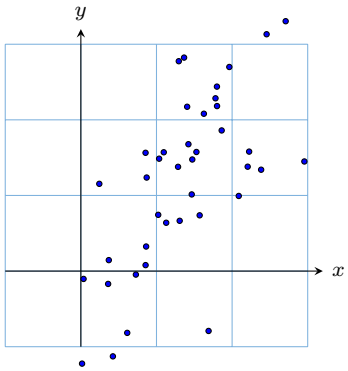
1. Explain whether this correlation implies that buying ice cream *causes* sharks to attack people.

2. Identify a likely confounding variable that influences both ice cream sales and the frequency of shark attacks.

E MEASURING LINEAR CORRELATION

E.1 ESTIMATING THE CORRELATION COEFFICIENT

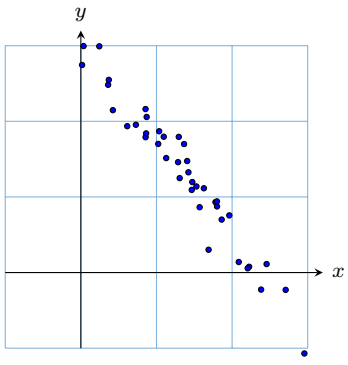
Ex 36: **Examine the scatter plot and choose the most appropriate correlation coefficient (r) from the options provided.**



The correlation coefficient is approximately:
- ☐ 1
- ☐ 0.9
- ☐ 0.7
- ☐ 0
- ☐ −0.7
- ☐ −0.9
- ☐ −1

.

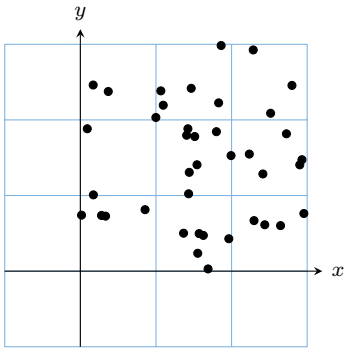
Ex 37: **Examine the scatter plot and choose the most appropriate correlation coefficient (r) from the options provided.**



The correlation coefficient is approximately:
- ☐ 1
- ☐ 0.9
- ☐ 0.7
- ☐ 0
- ☐ −0.7
- ☐ −0.9
- ☐ −1

.

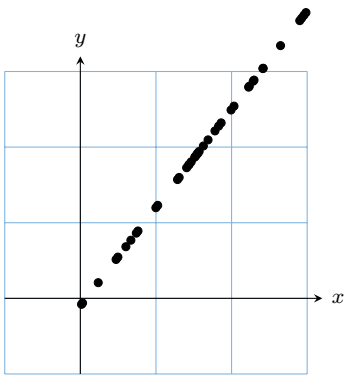
Ex 38: **Examine the scatter plot and choose the most appropriate correlation coefficient (r) from the options provided.**



The correlation coefficient is approximately:
- ☐ 1
- ☐ 0.9
- ☐ 0.7
- ☐ 0
- ☐ −0.7
- ☐ −0.9
- ☐ −1

.

Ex 39: **Examine the scatter plot and choose the most appropriate correlation coefficient (r) from the options provided.**



The correlation coefficient is approximately:
- ☐ 1
- ☐ 0.9
- ☐ 0.7
- ☐ 0
- ☐ −0.7
- ☐ −0.9
- ☐ −1

.

E.2 APPLYING CORRELATION ANALYSIS TO REAL-WORLD DATA

Ex 40: The following table shows the data for two quantitative variables, X and Y.

X	2	4	6	8	10
Y	5	12	18	26	31

1. Plot these points on a scatter diagram.
2. Calculate Pearson's correlation coefficient, r, with calculator.
3. Describe the correlation between variable X and variable Y in terms of direction and strength.

Ex 41: A shop owner records the amount spent on advertising (X, in hundreds of dollars) and the resulting sales revenue (Y, in thousands of dollars) for 6 consecutive months.

X	1	2	3	4	5	6
Y	10	18	25	40	55	62

1. Plot these points on a scatter diagram.
2. Calculate Pearson's correlation coefficient, r, with a calculator.
3. Describe the correlation between variable X and variable Y in terms of direction and strength.

Ex 42: A cafe manager tracks the daily average temperature (X in °C) and the number of hot chocolates sold (Y) over 6 days.

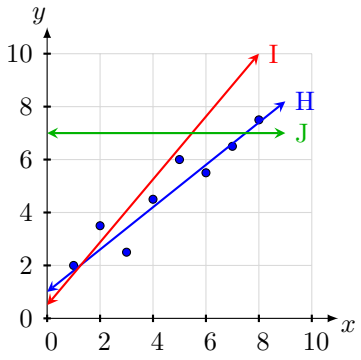
X	5	10	15	20	25	30
Y	150	120	90	60	40	25

1. Plot these points on a scatter diagram.
2. Calculate Pearson's correlation coefficient, r, with a calculator.
3. Describe the correlation between variable X and variable Y in terms of direction and strength.

F LINEAR REGRESSION

F.1 ESTIMATING THE LINE OF BEST FIT BY EYE

MCQ 43: Which of the following lines best represents the line of best fit for the data?



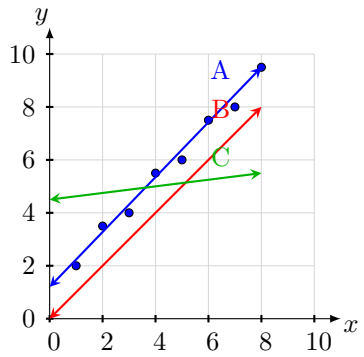
☐ H

☐ I

☐ J

☐ None of the lines fit the data well.

MCQ 44: Which of the following lines best represents the line of best fit for the data?



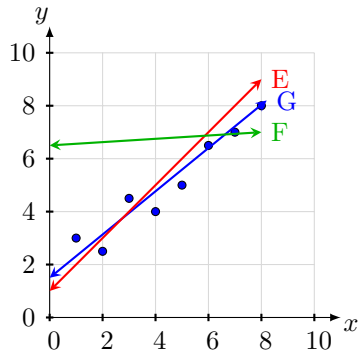
☐ A

☐ B

☐ C

☐ None of the lines fit the data well.

MCQ 45: Which of the following lines best represents the line of best fit for the data?



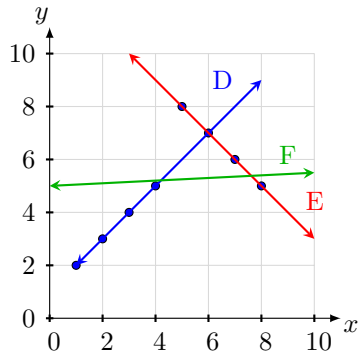
☐ E

☐ F

☐ G

☐ None of the lines fit the data well.

MCQ 46: Which of the following lines best represents the line of best fit for the data?



☐ D

☐ E

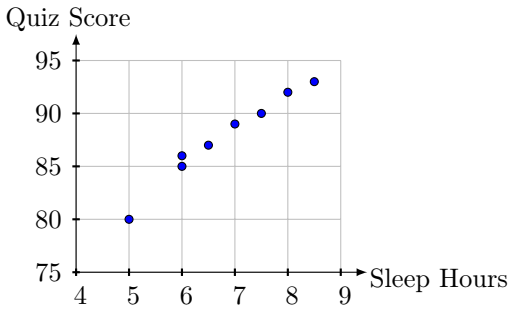
☐ F

☐ None of the lines fit the data well.

F.2 EVALUATING THE APPROPRIATENESS OF A LINEAR MODEL

Ex 47: Scenario: Lisa recorded her hours of sleep and quiz scores over 8 days.

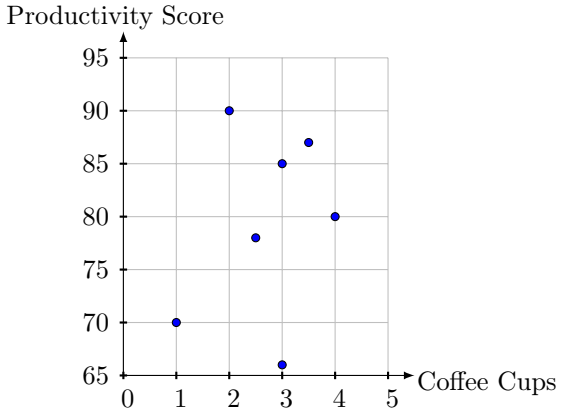
Sleep Hours	6	7.5	5	8	6.5	7	8.5	6
Quiz Score	85	90	80	92	87	89	93	86



Based on the scatter plot, is it reasonable to fit a line of best fit to this data?

MCQ 48: Scenario: John recorded his daily coffee consumption and a self-assessed productivity score.

Coffee Cups	1	3	2	4	2.5	3	3.5
Productivity Score	70	85	90	80	78	66	87



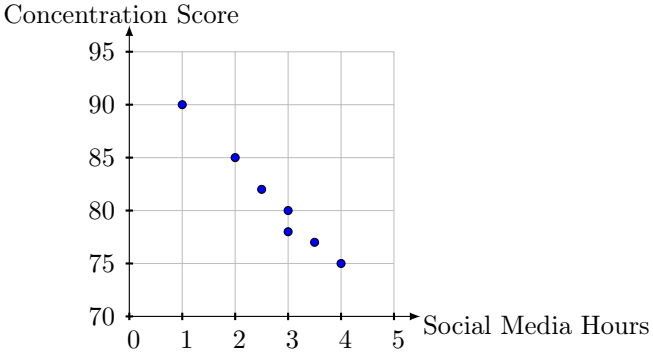
Is it reasonable to fit a line of best fit to this data?

☐ Yes

☐ No

MCQ 49: Scenario: Sophia recorded her daily social media hours and her ability to concentrate (scored out of 100).

Social Media Hours	1	3	2	4	2.5	3	3.5
Concentration Score	90	80	85	75	82	78	77



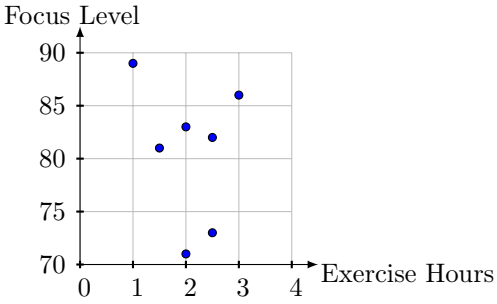
Is it reasonable to fit a line of best fit to this data?

☐ Yes

☐ No

MCQ 50: **Scenario:** Anna investigated if exercise impacts her mental focus, recording daily exercise hours and a focus score (out of 100).

Exercise Hours	1	2.5	2	3	1.5	2	2.5
Focus Level	89	82	71	86	81	83	73



Is it reasonable to fit a line of best fit to this data?

☐ Yes

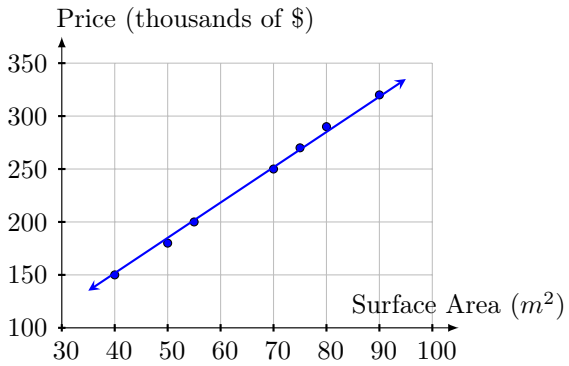
☐ No

F.3 ESTIMATING VALUES GRAPHICALLY

MCQ 51: Sophie, a real estate agent, explored the correlation between house surface area (in m^2) and price (in thousands of dollars). She recorded data from recent sales:

Surface Area (m^2)	40	50	55	70	75	80	90
Price (thousands of $)	150	180	200	250	270	290	320

She plotted the data and drew a line of best fit:



Using the line of best fit, estimate the price of a house with a surface area of 60 m^2.

☐ $165,000

☐ $200,000

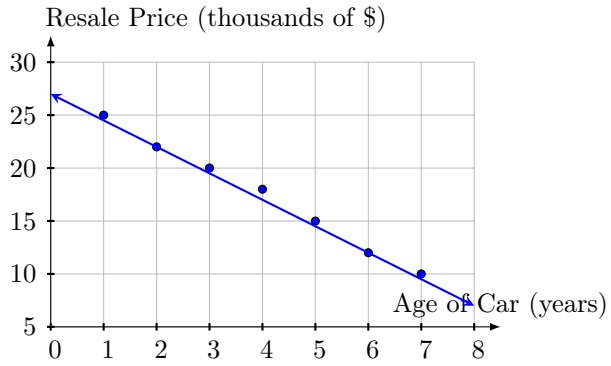
☐ $225,000

☐ $240,000

MCQ 52: Caroline, a used car dealer, investigated the correlation between car age (in years) and resale price (in thousands of dollars). She recorded data from recent sales:

Age of Car (years)	1	2	3	4	5	6	7
Resale Price (thousands of $)	25	22	20	18	15	12	10

She plotted the data and drew a line of best fit:



Using the line of best fit, estimate the resale price of a car that is 4.5 years old.

☐ $16,000

☐ $19,500

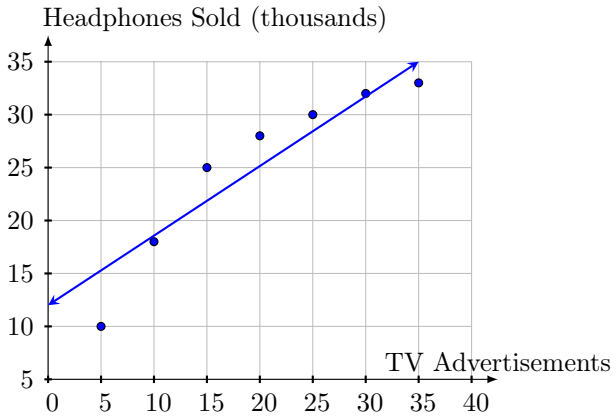
☐ $21,000

☐ $13,500

MCQ 53: Alex, a marketing manager, studied the correlation between TV advertisements aired weekly and headphone sales (in thousands). He recorded data over several weeks:

TV Advertisements	5	10	15	20	25	30	35
Headphones Sold (thousands)	10	18	25	28	30	32	33

He plotted the data and drew a line of best fit:



Using the line of best fit, estimate the number of headphones sold if 18 TV advertisements are aired in a week.

☐ 22,000

☐ 24,000

☐ 26,000

☐ 28,000

F.4 ESTIMATING VALUES WITH LINEAR EQUATIONS

Ex 54: Dr. Smith, a geneticist, studied the relationship between fathers' and sons' heights (in cm). He collected data from multiple families, where x represents the father's height and y represents the son's height. After analysis, he derived the best-fit line equation: $y = x + 2$. Using this equation, estimate the son's height if the father's height is 177 cm.

☐ cm

Ex 55: Ms. Lopez, a fitness coach, studied the relationship between weekly exercise hours and resting heart rates (in beats per minute, bpm) of her clients. She collected data from several clients, where x represents the hours of exercise per week and y represents the resting heart rate. After analysis, she derived the best-fit line equation: $y = -2x + 80$. Using this equation, estimate the resting heart rate for a client who exercises 6 hours per week.

☐ bpm

Ex 56: Mr. Patel, a nutritionist, investigated the relationship between daily water intake (in liters) and energy levels (on a scale of 0 to 10) among his clients. He collected data from several clients, where x represents the daily water intake and y represents the energy level. After analysis, he derived the best-fit line equation: $y = 1.8x + 2.5$. Using this equation, estimate the energy level for a client who drinks 3.5 liters of water daily.

☐

Ex 57: Ms. Chen, a teacher, explored the relationship between students' study time (in hours per week) and their test scores (out of 100). She collected data from her class, where x represents the study hours per week and y represents the test score. After analysis, she derived the best-fit line equation: $y = 4.2x + 55.6$. Using this equation, estimate the test score for a student who studies 8.5 hours per week.

☐

F.5 APPLYING LINEAR REGRESSION MODELS

Ex 58: A botanist measures the height (h) in cm and the number of leaves (L) of 8 young plants of the same species.

Height (h)	12	15	18	22	25	28	32	35
Leaves (L)	8	11	14	19	22	25	30	32

1. Write down the equation of the regression line of L on h in the form $L = ah + b$ using a calculator.
2. Interpret the value of a in the context of the problem.
3. Calculate the coefficient of determination, r^2, and explain what it indicates.
4. Estimate the number of leaves on a plant with a height of 20 cm.
5. Comment on the validity of using this model to estimate the number of leaves on a plant that is 60 cm tall.

Ex 59: A real estate agent records the living area (x) in square meters (m^2) and the selling price (y) in thousands of dollars of 6 recently sold apartments in a neighborhood.

Area (x)	60	75	90	110	130	150
Price (y)	180	220	250	310	350	410

1. Write down the equation of the regression line of y on x in the form $y = ax + b$ using a calculator.
2. Interpret the value of a in the context of the problem.
3. Calculate the coefficient of determination, r^2, and explain what it indicates.
4. Estimate the selling price of an apartment with an area of 100 m^2.
5. Comment on the validity of using this model to estimate the price of a castle with an area of 1000 m^2.

Ex 60: A car dealership tracks the age of a specific car model (x) in years and its resale value (V) in thousands of dollars.

Age (x)	1	2	3	5	6	8
Value (V)	25	22	19	14	12	7

1. Write down the equation of the regression line of V on x in the form $V = ax + b$ using a calculator.
2. Interpret the value of a in the context of the problem.
3. Calculate the coefficient of determination, r^2, and explain what it indicates.
4. Estimate the value of a car that is 4 years old.
5. Comment on the validity of using this model to estimate the value of a car that is 15 years old.