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)**
- "Height Increase (cm)" is the: **dependent variable (y)**

Answer: **Analysis:** We are investigating if sunlight *influences* growth. The change in height *depends on* the amount of sunlight.

- "Sunlight Hours per Day" is the **independent (explanatory) variable (x)** because it is the factor being changed or observed to see its effect.
- "Height Increase (cm)" is the **dependent (response) variable (y)** because it is the outcome that is measured.

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: **dependent variable (y)**
- "Exercise Hours per Week" is the: **independent variable (x)**

Answer: **Analysis:** We are investigating if exercise *impacts* heart rate. The heart rate is the outcome that *depends on* the amount of exercise.

- "Exercise Hours per Week" is the **independent (explanatory) variable (x)**.
- "Resting Heart Rate (bpm)" is the **dependent (response) variable (y)**.

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: **independent variable (x)**
- "Bread Height (cm)" is the: **dependent variable (y)**

Answer: **Analysis:** We are investigating if temperature *impacts* bread height. The bread's final height is the result that *depends on* the oven temperature.

- "Oven Temperature (°C)" is the **independent (explanatory) variable (x)**.
- "Bread Height (cm)" is the **dependent (response) variable (y)**.

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)**
- "Study Hours per Week" is the: **independent variable (x)**

Answer: **Analysis:** We are investigating if study time *influences* quiz scores. The score is the outcome that *depends on* the study time.

- "Study Hours per Week" is the **independent (explanatory) variable (x)**.
- "Quiz Score (out of 20)" is the **dependent (response) variable (y)**.

B SCATTER PLOTS

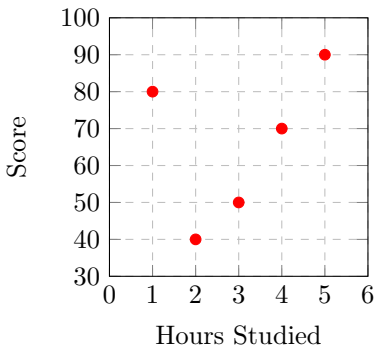
B.1 MATCHING DATA TABLES TO SCATTER PLOTS

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

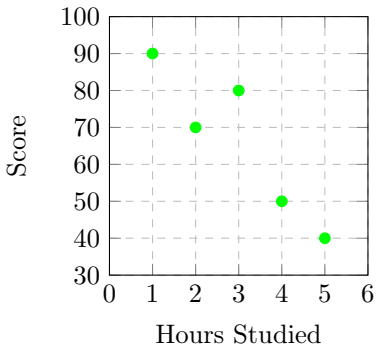
Which scatter plot correctly represents the data in the table?

Graph E



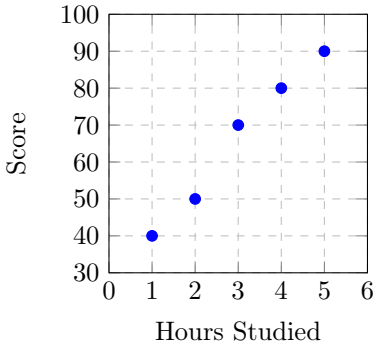
□

Graph F



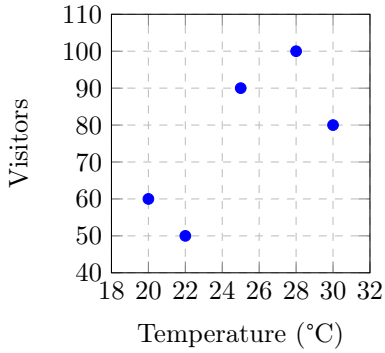
□

Graph G



☒

Graph C



☐

Answer: The correct answer is **Graph G**.
To verify, we can check a few key points from the table against the graphs:

- **Point O (1, 40):** Graph E has a point at (1, 80). Graph F has a point at (1, 90). Only Graph G has the correct point at (1, 40).
- **Point P (2, 50):** Graph E has a point at (2, 40). Graph F has a point at (2, 70). Only Graph G has the correct point at (2, 50).

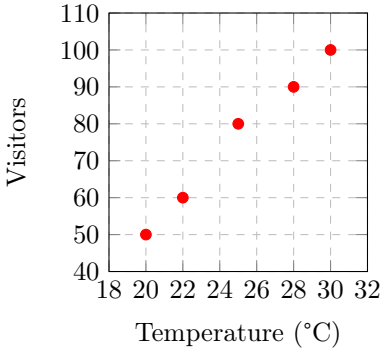
Based on these checks, only Graph G accurately plots all the data points from the table.

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

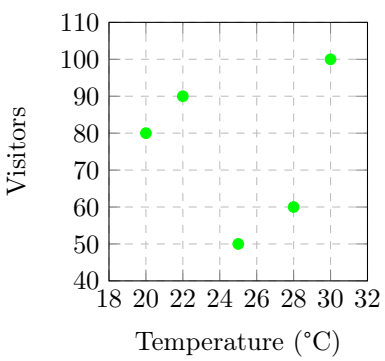
Which scatter plot correctly represents the data?

Graph A



☒

Graph B



☐

Answer: The correct answer is **Graph A**.
We can check the first point from the table, (20, 50):

- Graph A correctly plots a point at (20, 50).
- Graph B plots a point at (20, 80). This is incorrect.
- Graph C plots a point at (20, 60). This is incorrect.

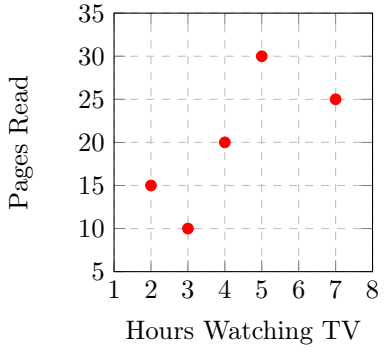
Only Graph A starts with the correct point. A further check of other points, like (30, 100), confirms that Graph A is the only correct representation.

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

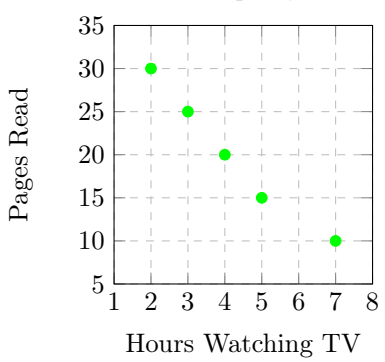
Which scatter plot correctly represents the data?

Graph P



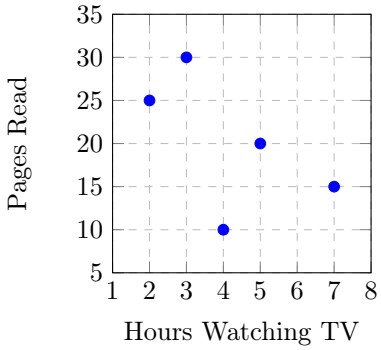
☐

Graph Q



☒

Graph R



□

Answer: The correct answer is **Graph Q**.
Let's check the point where TV hours (x) is highest: (7, 10).

- In Graph P, when $x = 7$, $y = 25$. This is incorrect.
- In Graph R, when $x = 7$, $y = 15$. This is incorrect.
- In Graph Q, when $x = 7$, $y = 10$. This is correct.

Checking another point, (2, 30), confirms that Graph Q is the only correct representation.

B.2 PLOTTING SCATTER PLOTS

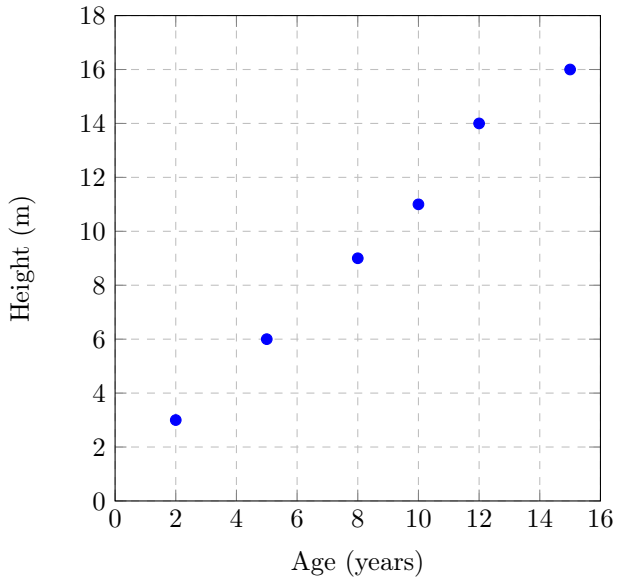
Ex 8: Scenario: A botanist measures the age and height of several trees of the same species. The collected data is shown in the table below.

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.

Answer: To create the scatter plot, we treat each tree's age as the independent variable (x) and its height as the dependent variable (y). We then plot the six coordinate pairs on the graph.
Completed Scatter Plot:

Tree Age vs. Height



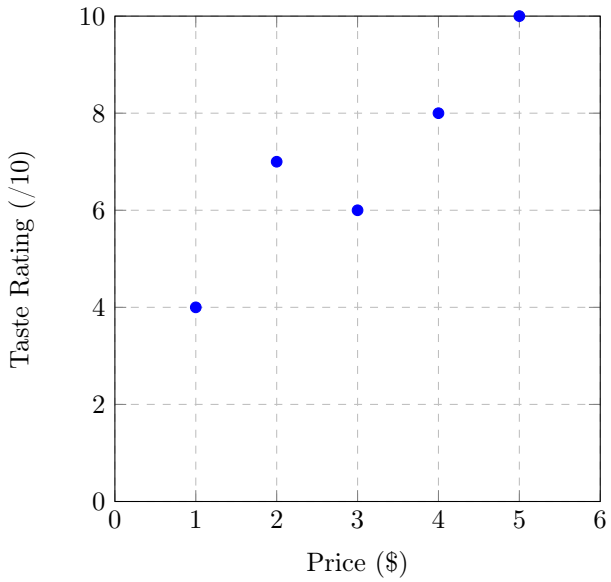
Ex 9: Scenario: A consumer group rates several brands of chocolate based on taste (out of 10) and records their price. The data is shown in the table below.

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.

Answer: To create the scatter plot, we treat the Price as the independent variable (x) and the Taste Rating as the dependent variable (y). We then plot the five coordinate pairs on the graph.
Completed Scatter Plot:

Chocolate Price vs. Taste Rating

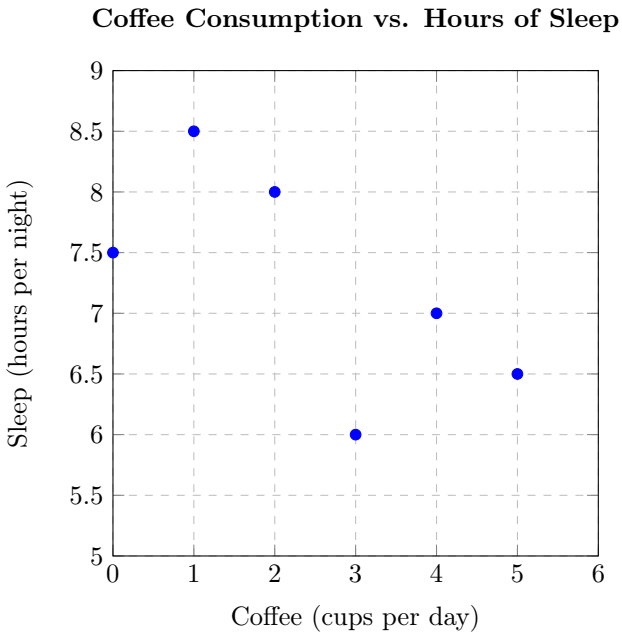


Ex 10: Scenario: A health researcher investigates a possible link between daily coffee consumption and hours of sleep. The data for six individuals is shown below.

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.

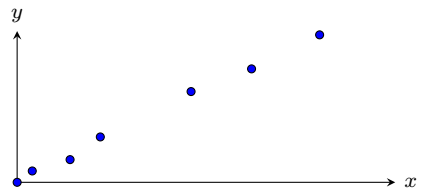
Answer: To create the scatter plot, we treat "Coffee (cups per day)" as the independent variable (x) and "Sleep (hours per night)" as the dependent variable (y). We then plot the six coordinate pairs on the graph.
Completed Scatter Plot:



C CORRELATION

C.1 IDENTIFYING THE DIRECTION OF A CORRELATION

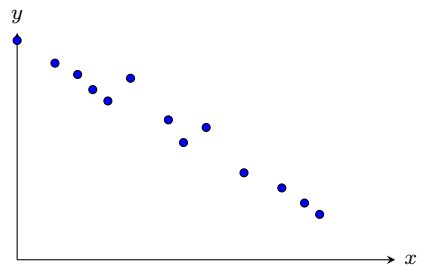
Ex 11: Determine the direction of the correlation shown in the scatter plot below.



The correlation is **positive**.

Answer: The correlation is **positive**. As the x values increase, the y values generally tend to increase, creating an upward trend from left to right.

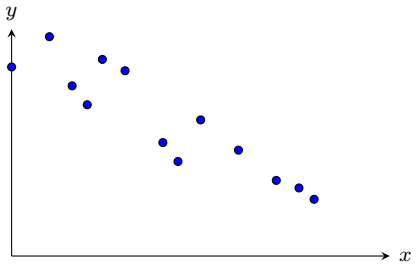
Ex 12: Determine the direction of the correlation shown in the scatter plot below.



The correlation is **negative**.

Answer: The correlation is **negative**. As the x values increase, the y values generally tend to decrease, creating a downward trend from left to right.

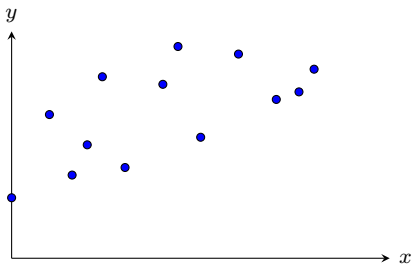
Ex 13: Determine the direction of the correlation shown in the scatter plot below.



The correlation is **negative**.

Answer: The correlation is **negative**. Although the points are scattered, the overall trend is downward from left to right.

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

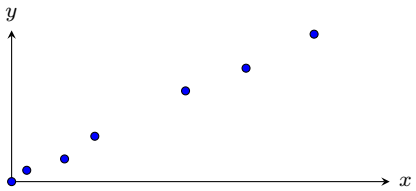


The correlation is **positive**.

Answer: The correlation is **positive**. Although the points are widely scattered, the overall trend is upward from left to right.

C.2 IDENTIFYING THE STRENGTH OF A CORRELATION

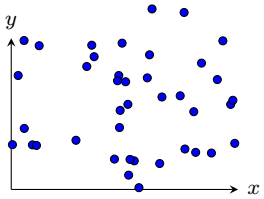
Ex 15: Determine the strength of the correlation shown in the scatter plot below.



The correlation strength is **Strong**.

Answer: The correlation strength is **strong**. The data points are clustered very tightly and form a clear, distinct linear pattern.

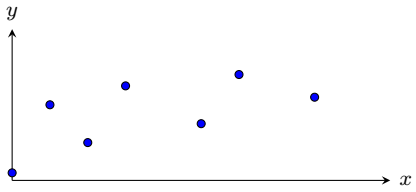
Ex 16: Determine the strength of the correlation shown in the scatter plot below.



The correlation strength is **None**.

Answer: There is **no** correlation. The points are widely and randomly scattered, showing no discernible upward or downward trend.

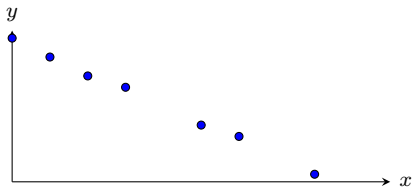
Ex 17: Determine the strength of the correlation shown in the scatter plot below.



The correlation strength is **Weak**.

Answer: The correlation strength is **weak**. While there is a slight upward (positive) trend, the points are very loosely scattered around a potential line.

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

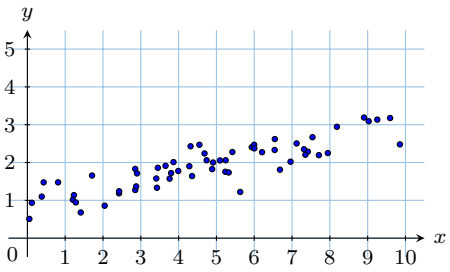


The correlation strength is **Strong**.

Answer: The correlation strength is **strong**. The data points are clustered tightly in a downward-sloping linear pattern.

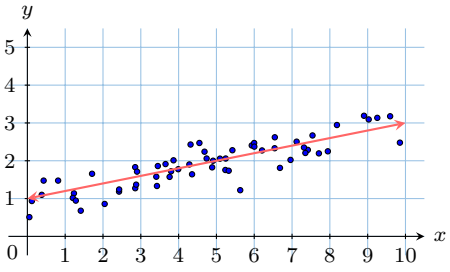
C.3 IDENTIFYING THE FORM OF A CORRELATION

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

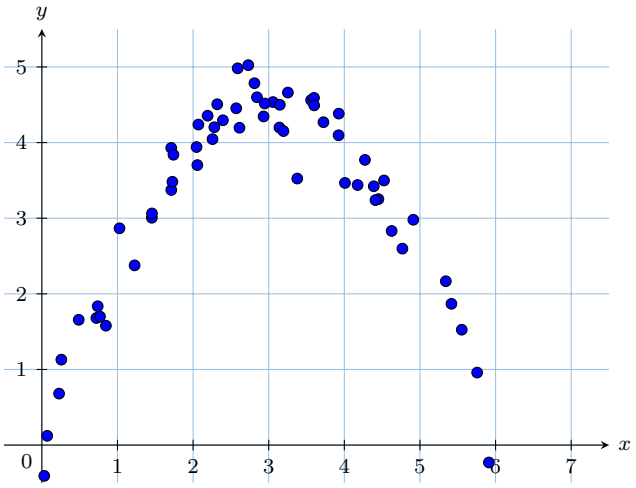


The correlation is **linear**.

Answer: The correlation is **linear** because the points generally follow a straight-line pattern.

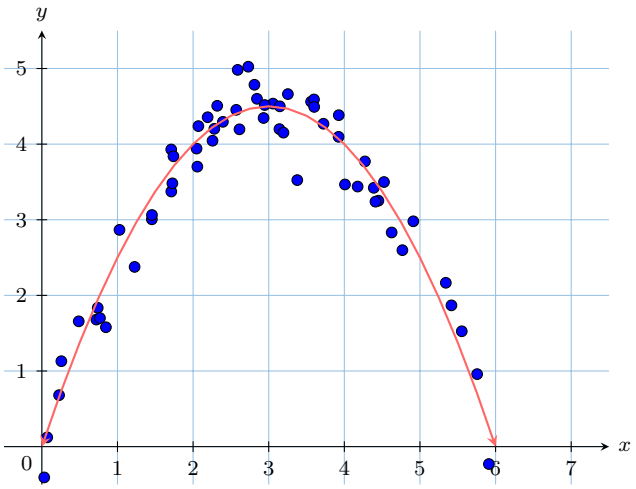


Ex 20: Determine if the form of the correlation is linear or non-linear.

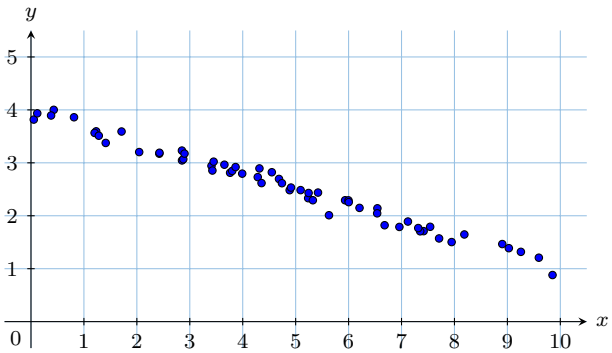


The correlation is **non-linear**.

Answer: The correlation is **non-linear** because the points clearly follow a curved (parabolic) pattern, not a straight line.

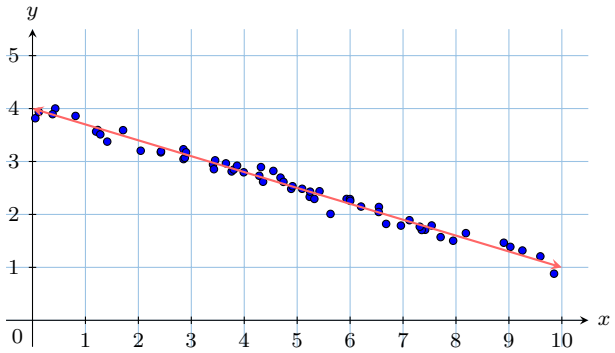


Ex 21: Determine if the form of the correlation is linear or non-linear.

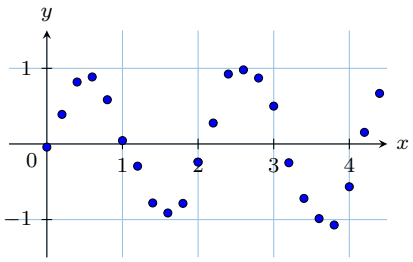


The correlation is **linear**.

Answer: The correlation is **linear** because the points generally follow a straight-line pattern.

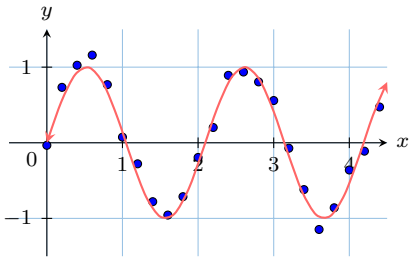


Ex 22: Determine if the form of the correlation is linear or non-linear.



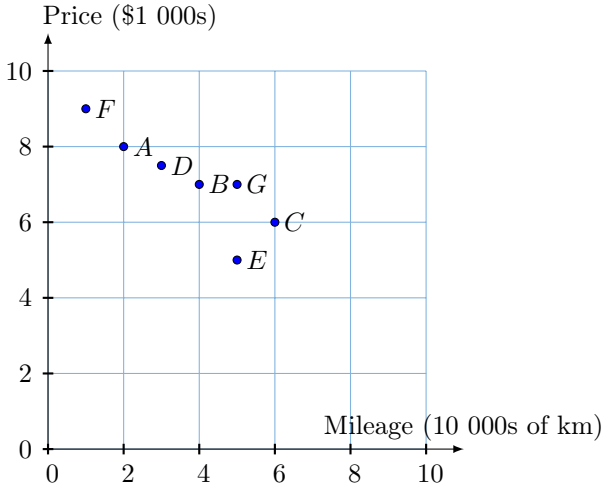
The correlation is **non-linear**.

Answer: The correlation is **non-linear**. The points follow a clear wave-like (sinusoidal) pattern, not a straight line.



C.4 INTERPRETING OUTLIERS IN CONTEXT

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

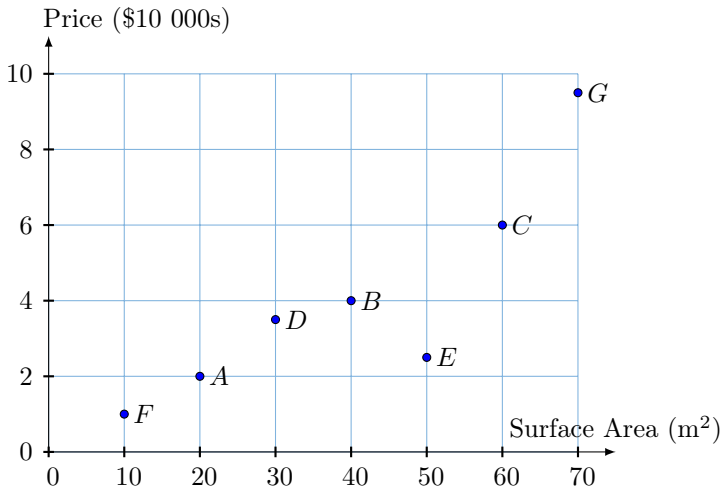


An outlier that represents a "good deal" would be a car that is underpriced for its mileage. Identify this car.
Select one: E.

Answer: The overall trend shows a negative correlation: as mileage increases, the price tends to decrease.

- An outlier representing a **good deal** will be a point significantly *below* this main trend.
- Point **E** (50,000 km, $5,000) is well below the other points with similar mileage (e.g., point G), indicating it is underpriced for its condition.
- Conversely, point G (50,000 km, $7,000) could be considered a "bad deal" as it is overpriced relative to the trend.

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

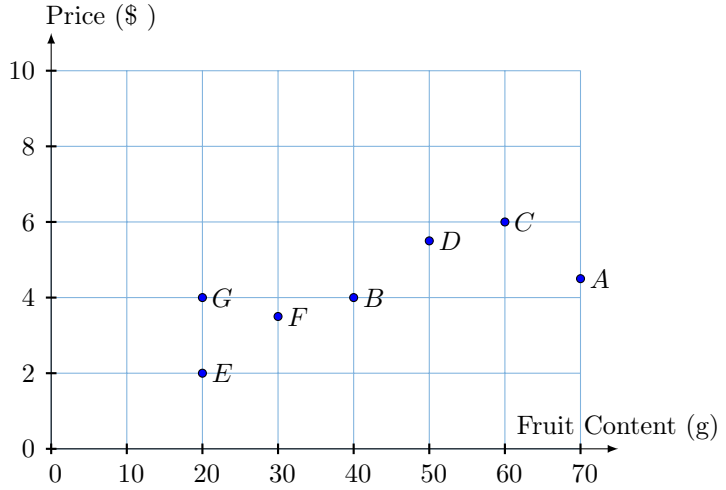


Which house is an outlier that represents a "good deal" (i.e., is underpriced for its size)?
Select one: E.

Answer: The overall trend shows a strong positive correlation: as the surface area increases, the price tends to increase.

- A "good deal" outlier will be a point significantly *below* the main trend.
- Point **E** (50 m², $25,000) deviates strongly from the trend. Other houses of similar size (e.g., 40m² for $40,000 or 60m² for $60,000) are much more expensive. This makes it a good deal.
- Point G (70 m², $95,000) is an outlier that could be a "bad deal" as it is priced higher than the trend would suggest.

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



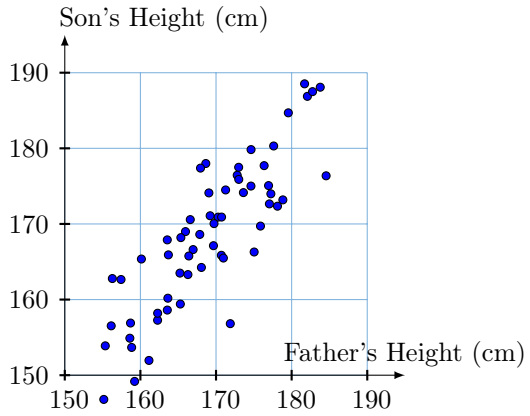
Which jam is an outlier that represents a "good deal" (a low price for a high fruit content)?
Select one: $\boxed{A}$.

Answer: The overall trend shows a positive correlation: as fruit content increases, the price tends to increase.

- A "good deal" outlier will have a high fruit content for a relatively low price, placing it below the main trend.
- Point **A** (70g, \$4.50) is a significant outlier. It has the highest fruit content but is priced lower than jams with less fruit (e.g., points C and D). This makes it a good deal.
- Point G (20g, \$4.00) is also an outlier, but it represents a "bad deal" as it has a low fruit content for a relatively high price.

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.

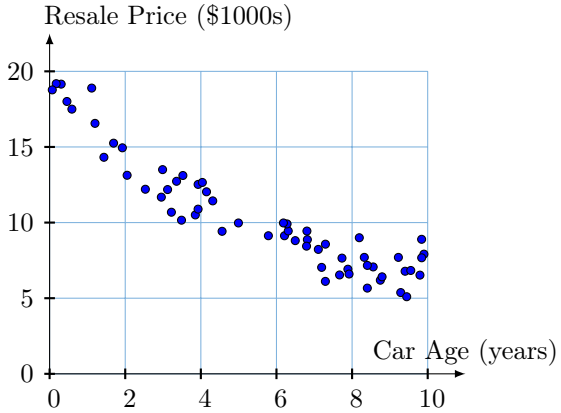


Which statement best describes the association shown in the plot?

- ☒ Taller fathers tend to have taller sons.
- ☐ Taller fathers tend to have shorter sons.
- ☐ There is no clear relationship between fathers' and sons' heights.

Answer: The correct statement is: **"Taller fathers tend to have taller sons."Justification:** The points on the scatter plot show a clear upward trend from left to right. This indicates a **positive correlation**, meaning that as the father's height (the independent variable) increases, the son's height (the dependent variable) also tends to increase.

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

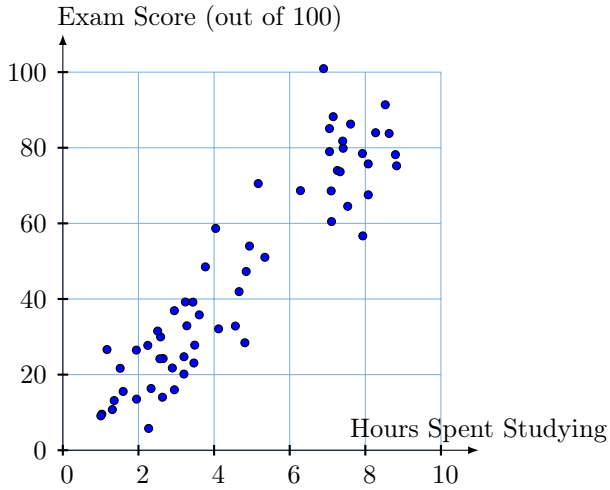


Which statement best describes the association shown in the plot?

- ☐ 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.

Answer: The correct statement is: **"Older cars tend to have lower resale prices."Justification:** The points on the scatter plot show a downward trend from left to right. This indicates a **negative correlation**, meaning that as the car's age (independent variable) increases, its resale price (dependent variable) tends to decrease.

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

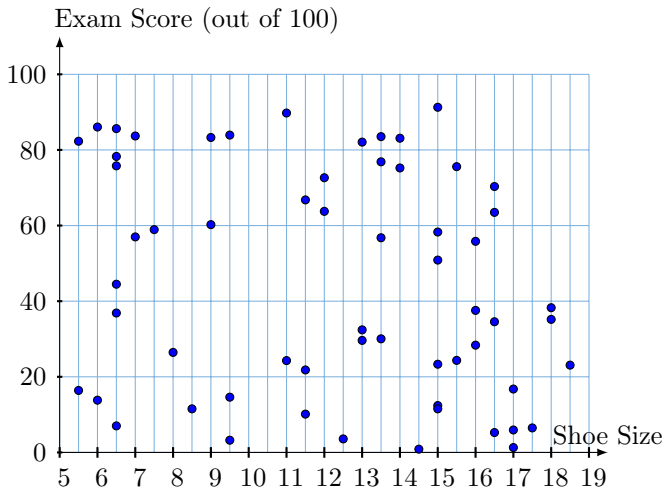


Which statement best describes the association shown in the plot?

- ☒ 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.

Answer: The correct statement is: **"Students who study more hours tend to score higher on the exam."Justification:** The points show a clear upward trend. This indicates a **positive correlation**, where an increase in the independent variable (study hours) is associated with an increase in the dependent variable (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.

Answer: The correct statement is: **"There is no clear relationship between shoe size and exam score."**
Justification: The points on the scatter plot are widely scattered with no discernible upward or downward trend. This indicates that there is **no correlation** between the two variables.

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: **Causation**
- Having energy and living a long life: **Correlation**

Answer:

- **Eating a nutritious diet and having energy: Causation.** There is a direct biological and scientifically established link where the nutrients from food are converted into energy for the body's cells. One directly causes the other.
- **Having energy and living a long life: Correlation.** While the two are associated, "having energy" does not directly cause a long life. Both could be the result of a confounding variable, such as good health or genetics.

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

Answer:

- **Sunny weather and eating ice cream: Causation.** Hot, sunny weather causes people to seek ways to cool down, which directly leads to an increase in ice cream consumption.
- **Eating ice cream and getting sunburned: Correlation.** These two events are correlated because they are both caused by the same confounding variable: **sunny weather**. Eating ice cream does not cause sunburn.
- **Sunny weather and getting sunburned: Causation.** There is a direct physical link. Exposure to ultraviolet (UV) radiation from the sun is the direct cause of sunburn.

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: **Causation**
- Getting good grades and wearing glasses: **Correlation**

Answer:

- **Studying hard and getting good grades: Causation.** There is a direct cause-and-effect link. Increased study time and effort (the cause) lead to better understanding and performance, which results in higher grades (the effect).
- **Getting good grades and wearing glasses: Correlation.** These two are often correlated, but there is no causal link. A confounding variable could be "time spent reading." A student who reads a lot may get good grades and also develop a need for glasses due to eye strain. One does not cause the other.

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.

Answer:

1. **Correlation vs Causation:** No, correlation does not imply causation. Just because the variables increase together does not mean firefighters cause the damage.
2. **Confounding Variable:** A likely confounding variable is the **size/severity of the fire**. A larger fire causes more damage (y) AND requires more firefighters to be deployed (x).

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.

Answer:

1. **Correlation vs Causation:** No, having big feet does not make you read better. The physical size of a foot has no biological effect on the brain's ability to read.
2. **Confounding Variable:** The confounding variable is the **age** of the child. As children get older, their feet grow larger (x), AND they receive more education, improving their reading scores (y).

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.

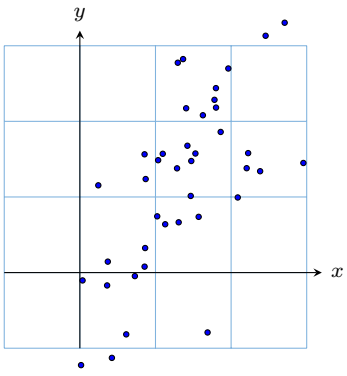
Answer:

1. **Correlation vs Causation:** No, eating ice cream does not cause shark attacks. The sharks are not attracted to the ice cream on the beach.
2. **Confounding Variable:** The confounding variable is the **temperature** (or season). When it is hot/summer, more people buy ice cream (x), AND more people go swimming in the ocean, increasing the risk of shark encounters (y).

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.



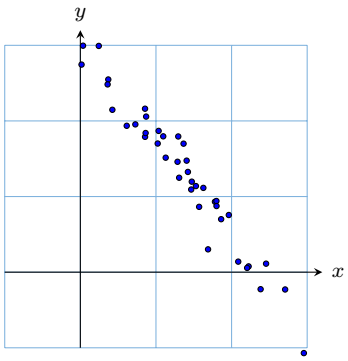
The correlation coefficient is approximately: $\boxed{0.7}$.

Answer: The scatter plot shows a **positive** correlation, so we can eliminate the negative options. The points show a clear linear trend but are somewhat scattered, not forming a tight line.

- $r = 1$ and $r = 0.9$ would imply a stronger, tighter pattern.
- $r = 0$ would imply no pattern.

Therefore, $r \approx 0.7$ is the most appropriate choice, representing a moderate to strong positive linear correlation.

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



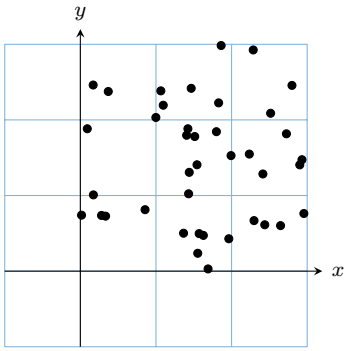
The correlation coefficient is approximately: $\boxed{-0.9}$.

Answer: The scatter plot shows a **negative** correlation, so we can eliminate the positive options. The points are clustered very tightly in a downward linear pattern.

- $r = -1$ would imply a perfect straight line.
- $r = -0.7$ would imply more scatter.

Therefore, $r \approx -0.9$ is the most appropriate choice, representing a very strong negative linear correlation.

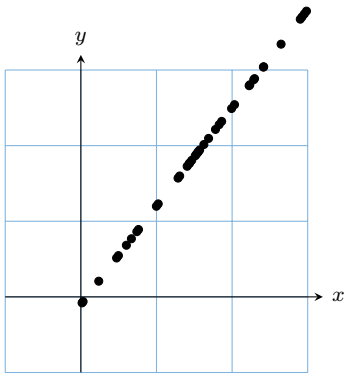
Ex 38: Examine the scatter plot and choose the most appropriate correlation coefficient (r) from the options provided.



The correlation coefficient is approximately: $\boxed{0}$.

Answer: There is **no discernible linear trend** in the data. The points are scattered randomly across the plot without any clear upward or downward direction. Therefore, the most appropriate choice is $r \approx 0$.

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



The correlation coefficient is approximately: $\boxed{1}$.

Answer: The scatter plot shows a **perfect positive linear correlation**. All the data points lie exactly on a single straight line that trends upward. Therefore, the correct correlation coefficient is $r = 1$.

E.2 APPLYING CORRELATION ANALYSIS TO REAL-WORLD DATA

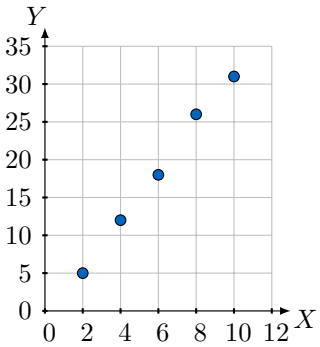
Ex 40: The following table shows the data for two quantitative variables, X and Y.

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.

Answer:

1. **Scatter Diagram:**



2. **Calculate r:** Using technology (LinReg), $r \approx 0.998$.
3. **Description:** There is a **very strong, positive linear correlation** between X and Y.

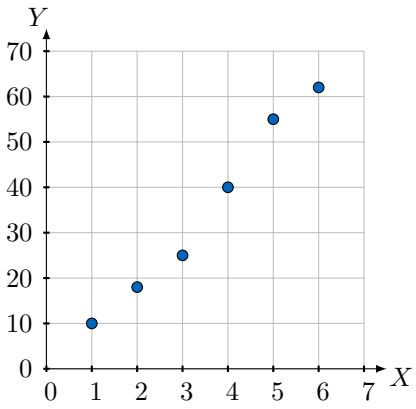
Ex 41: A shop owner records the amount spent on advertising (X, in hundreds of dollars) and the resulting sales revenue (Y, in thousands of dollars) for 6 consecutive months.

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.

Answer:

1. **Scatter Diagram:**



2. **Calculate r:** Using technology, $r \approx 0.985$.
3. **Description:** There is a **very strong, positive linear correlation** between advertising spend and sales revenue.

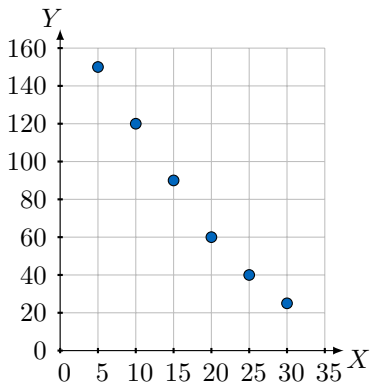
Ex 42: A cafe manager tracks the daily average temperature (X in °C) and the number of hot chocolates sold (Y) over 6 days.

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.

Answer:

1. **Scatter Diagram:**

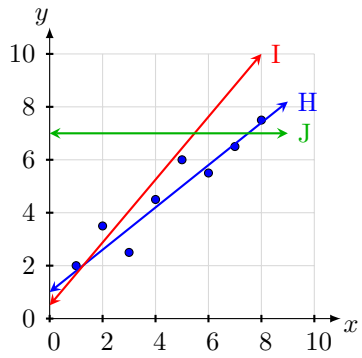


2. **Calculate r:** Using technology, $r \approx -0.996$.
3. **Description:** There is a **very strong, negative linear correlation** between temperature and hot chocolate sales.

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

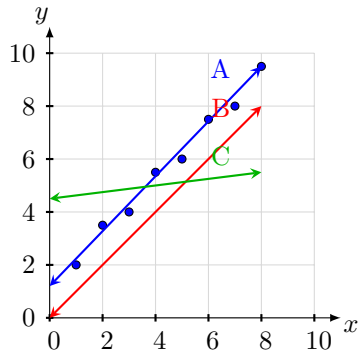
☐ J

☐ None of the lines fit the data well.

Answer: The line of best fit should capture the central, upward trend of the data.

- **Line H (blue)** is the best fit. It follows the positive slope of the data and passes through the middle of the point cloud, balancing the number of points above and below it.
- **Line I (red)** is too steep and does not accurately represent the slope of the data.
- **Line J (green)** is too flat and positioned too high, completely missing the data's trend.

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



☒ A

☐ B

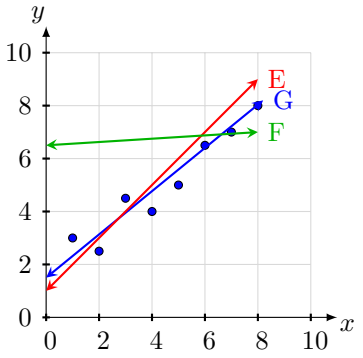
☐ C

☐ None of the lines fit the data well.

Answer: The data shows a strong, positive linear correlation.

- **Line A (blue)** is the best fit. It accurately captures both the steep positive slope and the central path of the data points.
- **Line B (red)** has a similar slope but is positioned too low, with most of the data points lying above it.
- **Line C (green)** is far too flat and does not represent the strong upward trend of the data.

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



☐ E

☐ F

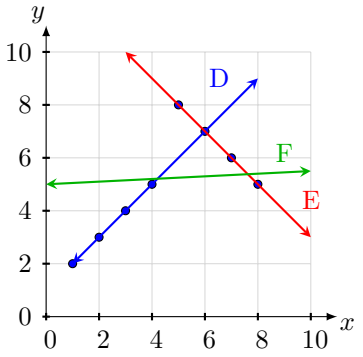
☒ G

☐ None of the lines fit the data well.

Answer: The data shows a moderate, positive linear correlation.

- **Line G (blue)** is the best fit. It effectively follows the upward trend and is positioned centrally within the scattered points.
- **Line E (red)** has a slightly steeper slope than the data trend suggests.
- **Line F (green)** is nearly horizontal and fails to capture the positive correlation in the data.

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



☐ D

☐ E

☐ F

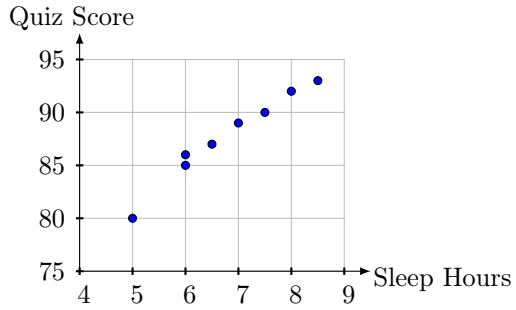
☒ None of the lines fit the data well.

Answer: **None of the lines fit the data well.**
Justification: The data is clearly **non-linear**. It follows a curved, inverted V-shape pattern. A line of best fit is only appropriate for data that shows a linear trend. None of the straight lines can adequately model this relationship.

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

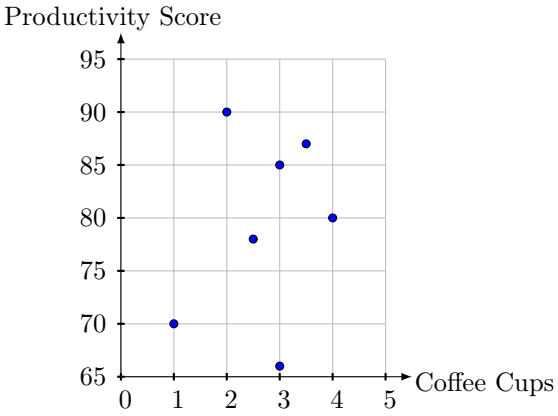


Based on the scatter plot, is it reasonable to fit a line of best fit to this data?

Answer: **Yes**. The scatter plot shows a clear **positive linear correlation**. As the number of sleep hours increases, the quiz scores tend to increase in a pattern that can be reasonably approximated by a straight line.

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?

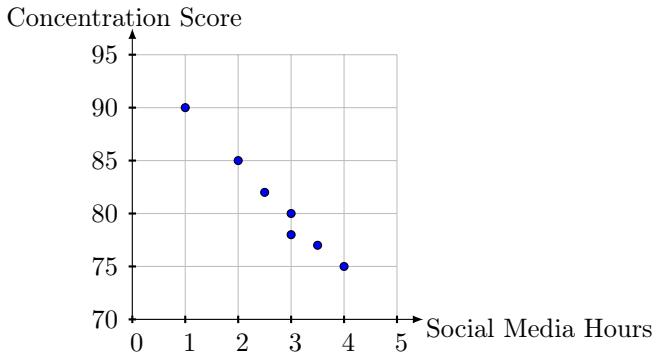
☐ Yes

☒ No

Answer: **No**. The points on the scatter plot are widely scattered and do not show a clear linear pattern. For example, a consumption of 3 cups is associated with scores as low as 66 and as high as 85. Since there is **no discernible correlation**, fitting a linear model would not be reasonable or useful.

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?

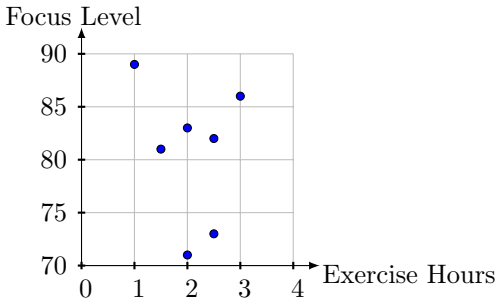
☒ Yes

☐ No

Answer: **Yes**. The scatter plot shows a fairly clear **negative linear correlation**. As the hours spent on social media increase, the concentration score tends to decrease. This consistent downward trend makes it reasonable to model the relationship with a line of best fit.

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

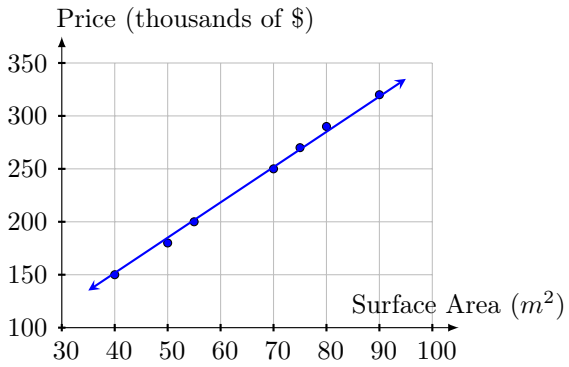
Answer: **No**. The points are widely scattered and do not suggest any clear pattern, either positive or negative. For example, 2.5 hours of exercise is associated with focus levels of both 73 and 82. Since there is **no clear correlation**, fitting a linear model would not be appropriate.

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

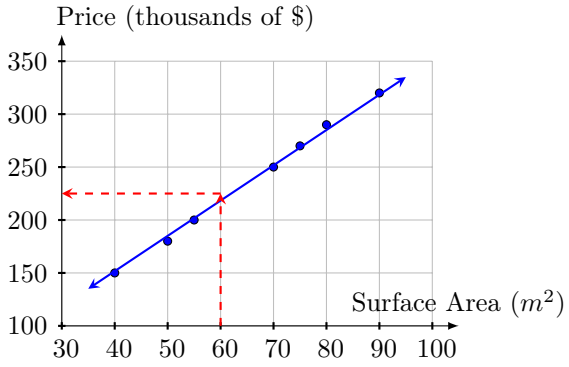
☒ $225,000

☐ $240,000

Answer: **Analysis:** To estimate the price for a 60 m^2 house, we follow these steps:

1. Find the value **60** on the horizontal axis (Surface Area).
2. Move vertically upwards from this point to the line of best fit.
3. From that point on the line, move horizontally to the left to read the corresponding value on the vertical axis (Price).

This process, shown by the red dashed lines on the graph below, indicates a value of approximately 225 on the vertical axis.

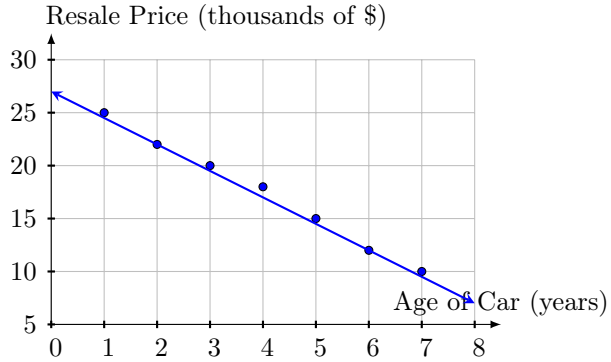


Therefore, the estimated price is $**225,000**.

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

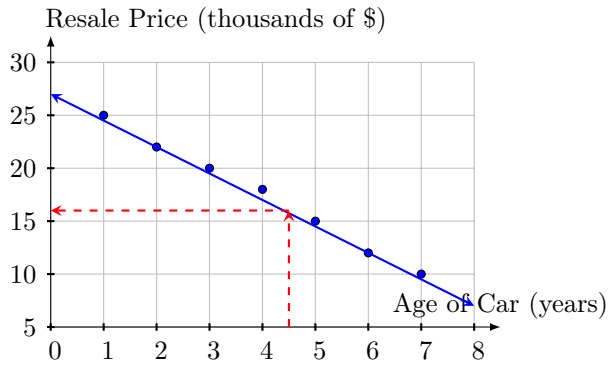
☐ $21,000

☐ $13,500

Answer: **Analysis:** To estimate the price for a 4.5-year-old car, we use the graph:

1. Find the value **4.5** on the horizontal axis (Age of Car).
2. Move vertically upwards from 4.5 to the line of best fit.
3. From that point, move horizontally to the left to find the value on the vertical axis (Resale Price).

As shown by the red dashed lines on the graph, this value is approximately 16.

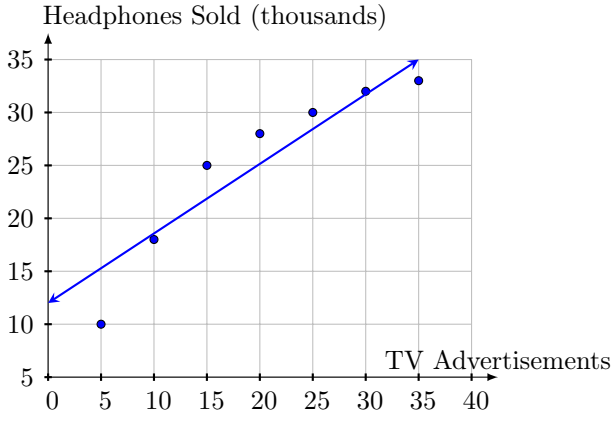


Thus, the estimated resale price is $**16,000**.

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

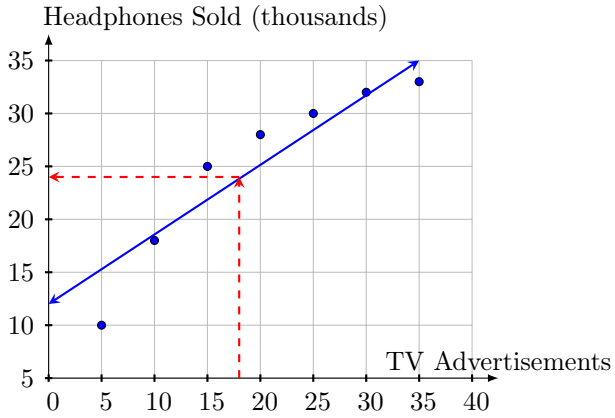
☐ 26,000

☐ 28,000

Answer: **Analysis:** To estimate the sales for 18 TV advertisements, we perform the following steps:

1. Find the value **18** on the horizontal axis (TV Advertisements).
2. Trace vertically up from 18 to the line of best fit.
3. From that point on the line, trace horizontally to the left to read the value on the vertical axis (Headphones Sold).

This procedure, illustrated by the red dashed lines, points to a value of approximately 24 on the vertical axis.



Therefore, the estimated number of headphones sold is **24,000**.

F.4 ESTIMATING VALUES WITH LINEAR EQUATIONS

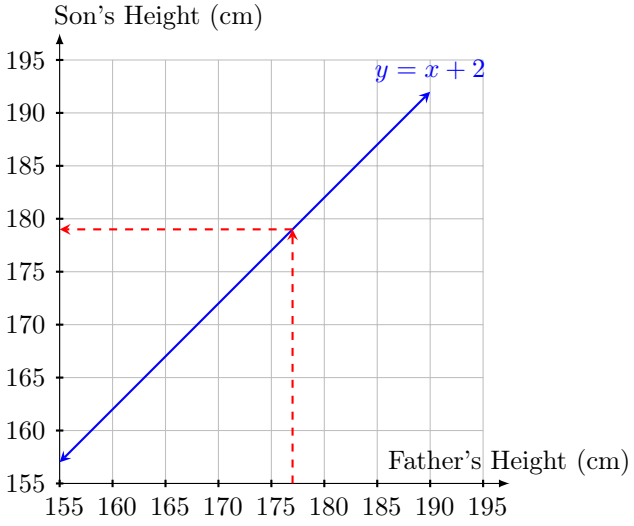
Ex 54: Dr. Smith, a geneticist, studied the relationship between fathers' and sons' heights (in cm). He collected data from multiple families, where x represents the father's height and y represents the son's height. After analysis, he derived the best-fit line equation: $y = x + 2$. Using this equation,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

$\boxed{179}$ cm

Answer: **Solution:**

1. **Identify the given information:**
   - The equation of the best-fit line is $y = x + 2$.
   - The father's height is given as $x = 177$ cm.
2. **Substitute the value of x into the equation:**
   To find the estimated son's height (y), we replace x with 177:
   $$y = 177 + 2$$
3. **Calculate the result:**
   $$y = 179$$

Therefore, the estimated height for the son is **179 cm**.
Visual Confirmation:
The graph below illustrates how substituting $x = 177$ into the equation yields a corresponding y-value of 179.



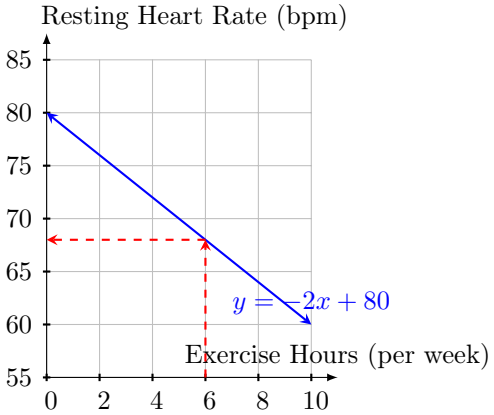
Ex 55: Ms. Lopez, a fitness coach, studied the relationship between weekly exercise hours and resting heart rates (in beats per minute, bpm) of her clients. She collected data from several clients, where x represents the hours of exercise per week and y represents the resting heart rate. After analysis, 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.

$\boxed{68}$ bpm

Answer: **Solution:**

1. **Identify the given information:**
   - The equation of the best-fit line is $y = -2x + 80$.
   - The weekly exercise hours are given as $x = 6$.
2. **Substitute the value of x into the equation:**
   To find the estimated resting heart rate (y), we replace x with 6:
   $$y = -2(6) + 80$$
3. **Calculate the result:**
   $$y = -12 + 80 = 68$$

Therefore, the estimated resting heart rate is **68 bpm**.
Visual Confirmation:
The graph below shows that for $x = 6$ hours of exercise, the corresponding estimated heart rate on the line is $y = 68$ bpm.



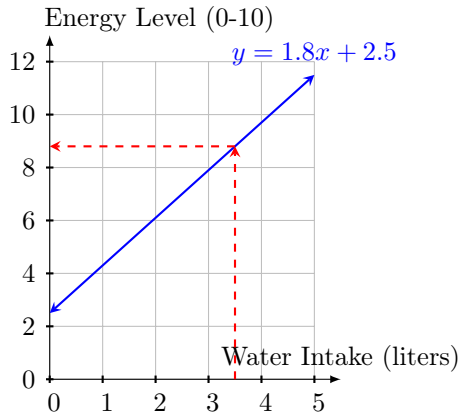
Ex 56: Mr. Patel, a nutritionist, investigated the relationship between daily water intake (in liters) and energy levels (on a scale of 0 to 10) among his clients. He collected data from several clients, where x represents the daily water intake and y represents the energy level. After analysis, he derived 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.

$$\boxed{8.8}$$

Answer: **Solution:**

1. **Identify the given information:**
   - The equation of the best-fit line is $y = 1.8x + 2.5$.
   - The daily water intake is given as $x = 3.5$ liters.
2. **Substitute the value of x into the equation:**
   To find the estimated energy level (y), we replace x with 3.5:
   $$y = 1.8(3.5) + 2.5$$
3. **Calculate the result:**
   $$y = 6.3 + 2.5 = 8.8$$

Therefore, the estimated energy level is **8.8**.
Visual Confirmation:
The graph below confirms that for a water intake of $x = 3.5$ liters, the estimated energy level is $y = 8.8$.



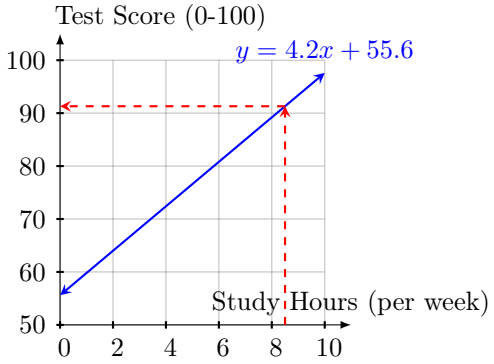
Ex 57: Ms. Chen, a teacher, explored the relationship between students' study time (in hours per week) and their test scores (out of 100). She collected data from her class, where x represents the study hours per week and y represents the test score. After analysis, she derived the best-fit line equation: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

$$\boxed{91.3}$$

Answer: **Solution:**

1. **Identify the given information:**
   - The equation of the best-fit line is $y = 4.2x + 55.6$.
   - The weekly study time is given as $x = 8.5$ hours.
2. **Substitute the value of x into the equation:**
   To find the estimated test score (y), we replace x with 8.5:
   $$y = 4.2(8.5) + 55.6$$
3. **Calculate the result:**
   $$y = 35.7 + 55.6 = 91.3$$

Therefore, the estimated test score is **91.3**.
Visual Confirmation:
The graph below shows that for $x = 8.5$ study hours, the corresponding estimated score on the line is $y = 91.3$.



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.

Answer:

1. **Regression Equation:** Using a calculator, $L = 1.07h - 5.05$.
2. **Interpretation of a:** The gradient $a \approx 1.07$ means that for every 1 cm increase in plant height, the number of leaves increases by approximately 1.07 leaves.
3. **Coefficient of Determination:** $r^2 \approx 0.995$. This indicates that 99.5% of the variation in the number of leaves can be explained by the height of the plant.
4. **Estimation ($h = 20$):** $L = 1.07(20) - 5.05 = 16.35$. Approximately 16 leaves.

5. **Validity ($h = 60$):** This estimate is **not valid** (unreliable) because $h = 60$ is far outside the range of the original data ($12 \leq h \leq 35$). This is **extrapolation**, and the linear growth pattern may not continue (the plant might stop growing leaves).

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.

Answer:

1. **Regression Equation:** Using a calculator, $y = 2.51x + 29.7$.
2. **Interpretation of a:** The gradient $a \approx 2.51$ means that for every additional square meter of living area, the selling price increases by approximately 2.51 thousand dollars ($2510).
3. **Coefficient of Determination:** $r^2 \approx 0.996$. This indicates that 99.6% of the variation in the selling price can be explained by the variation in the living area.
4. **Estimation ($x = 100$):** $y = 2.51(100)+29.7 = 251+29.7 = 280.7$. The estimated price is $\approx \$280,700$.
5. **Validity ($x = 1000$):** This estimate is **not valid** (unreliable) because $x = 1000$ is far outside the range of the collected data ($60 \leq x \leq 150$). This is **extrapolation**, and the pricing model for standard apartments likely does not apply to castles.

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.

Answer:

1. **Regression Equation:** Using a calculator, $V = -2.53x + 27.3$.
2. **Interpretation of a:** The gradient $a \approx -2.53$ means that for every additional year of age, the car's value decreases (depreciates) by approximately 2.53 thousand dollars ($2530).
3. **Coefficient of Determination:** $r^2 \approx 0.993$. This indicates that 99.3% of the variation in the car's value can be explained by its age.
4. **Estimation ($x = 4$):** $V = -2.53(4) + 27.3 = -10.12 + 27.3 = 17.18$. Approximately $17,180.
5. **Validity ($x = 15$):** This estimate is **not valid** (unreliable). It is extrapolation. Mathematically, $V = -2.53(15)+27.3 = -10.65$, which predicts a negative value for the car, which is impossible.