Name: _____ Date: _____

UNIT 5

4.5.7 — Correlation

Bivariate data • direction • strength • context

A. CORRELATION FUNDAMENTALS

1. Complete each statement: positive correlation means that as one variable increases, the other tends to **_____**; negative correlation means it tends to **_____**.
2. Does a strong correlation prove that one variable causes the other? **Answer =** _____
3. Why is the phrase **"tends to"** useful when describing correlation?

B. READ SCATTER DIAGRAMS

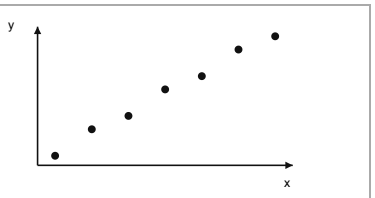


Diagram 1

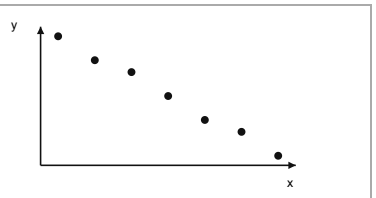


Diagram 2

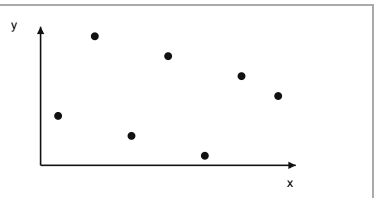


Diagram 3

1. Describe the **direction and strength** of each diagram.
2. Which diagram shows the strongest linear association? **Diagram =** _____

C. WRITE CONCLUSIONS IN CONTEXT

1. A scatter diagram of study time and test score shows a fairly strong positive correlation. Write a cautious conclusion in context.
2. A scatter diagram of outside temperature and heating cost shows a strong negative correlation. Write a cautious conclusion in context.

Name: _____ Date: _____

UNIT 5

4.5.7 — Correlation

Outliers • third variables • causation

D. AN UNUSUAL POINT

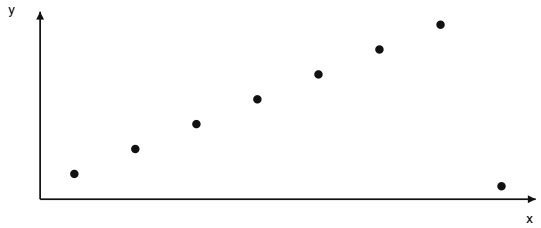


Diagram: Most points follow an upward trend, but one point lies far away from the pattern.

1. Without the unusual point, what kind of correlation would you expect? **Correlation =** _____
2. How can the unusual point affect the apparent strength of the correlation?
3. Should the point automatically be deleted? Explain.

E. CORRELATION IS NOT CAUSATION

1. Ice-cream sales and sunburn cases are positively correlated. Suggest a third variable that could help explain both.
2. A survey finds that students who own more books tend to score higher in reading. Give one reason why this does **not** prove that buying books causes higher scores.
3. Rewrite the claim “More exercise causes better sleep” as a cautious correlation statement.

F. FIND AND CORRECT THE ERROR

1. A student says, “The points go downward, so the correlation is strong negative.” What information is missing from this reasoning?
2. A student says, “No linear correlation means the variables have no relationship of any kind.” Explain why this can be wrong.

Name: _____ Date: _____

UNIT 5

4.5.7 — Correlation

IB-style extension • Pearson r • linearity

IB-STYLE EXTENSION BEYOND SOURCE DECK

G. PEARSON'S CORRELATION COEFFICIENT, R

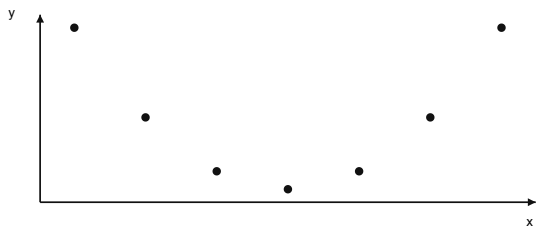
Use technology-produced values of *r*. Focus on sign, closeness to ±1, and whether a linear model is appropriate.

1. For each value, state the direction of linear correlation: **r = 0.91**, **r = −0.74**, **r = 0.08**.
2. Which shows the strongest linear association: **r = 0.62** or **r = −0.88**? Explain.

3. A calculator gives **r = −0.96**. Write a suitable interpretation without claiming causation.

IB-STYLE EXTENSION BEYOND SOURCE DECK

H. R MEASURES LINEAR ASSOCIATION



The points follow a clear curved pattern. For this symmetric data set, Pearson's *r* is approximately 0.

1. Why would it be misleading to say "there is no relationship" just because *r* is near 0?

2. What kind of statement is safer? **There is little or no _____ correlation.** _____

I. COMPARE EVIDENCE

1. Data set A has $r = 0.97$ and data set B has $r = 0.41$. Which has the stronger **linear** association?

   Set = _____

2. Can you decide from these two *r*-values which relationship is more important in the real world? Explain.

3. Give one reason a single outlier may change *r* substantially.

Name: _____ Date: _____

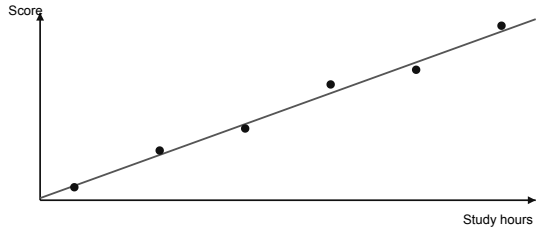
UNIT 5

4.5.7 — Correlation

IB-style extension • line of best fit • prediction

IB-STYLE EXTENSION BEYOND SOURCE DECK

J. REGRESSION LINE IN CONTEXT



A model for predicted test score is **y = 4.2x + 38**, where *x* is study hours and *y* is predicted score.

1. Interpret the gradient 4.2 in context.

2. Predict the score for a student who studies 5 hours. **Score =** _____

3. What does the intercept 38 represent in this model?

IB-STYLE EXTENSION BEYOND SOURCE DECK

K. INTERPOLATION AND EXTRAPOLATION

1. The observed study times range from 1 to 6 hours. Is a prediction for 4.5 hours interpolation or extrapolation?

   Type = _____

2. Is a prediction for 12 hours interpolation or extrapolation? **Type =** _____

3. Explain why the 12-hour prediction should be treated cautiously.

L. CHALLENGE — EVALUATE A CLAIM

1. A report states: "Because $r = 0.93$ between attendance and exam score, increasing attendance will definitely raise every student's score." Identify **two** weaknesses in this claim.

2. A second study finds $r = 0.18$ but the scatter plot shows a clear U-shaped pattern. Explain why quoting r alone is inadequate.

3. Write a three-part checklist for interpreting bivariate data responsibly.

ANSWER KEY

Name: _____ Date: _____

UNIT 5

4.5.7 — Correlation

ANSWER KEY

A. CORRELATION FUNDAMENTALS

1. Increase; decrease.
2. No.
3. Because individual points do not have to follow the pattern exactly; correlation describes a tendency across the data.

B. READ SCATTER DIAGRAMS

1. Diagram 1: strong positive. Diagram 2: strong negative. Diagram 3: little or no linear correlation.
2. Diagrams 1 and 2 are both extremely strong; their strengths are approximately equal, but their directions are opposite.

C. WRITE CONCLUSIONS IN CONTEXT

1. Students who study for more hours tend to achieve higher test scores.
2. As outside temperature increases, heating cost tends to decrease.

D. AN UNUSUAL POINT

1. A strong positive correlation.
2. It can weaken the apparent linear association because it lies far from the main trend.
3. No. First check whether it is a valid observation, a recording error, or a genuinely unusual case. Deleting valid data without justification can distort conclusions.

Name: _____ Date: _____

UNIT 5

4.5.7 — Correlation

ANSWER KEY

E. CORRELATION IS NOT CAUSATION

1. Temperature or sunshine could increase both ice-cream sales and sunburn cases.
2. Other variables such as home learning environment, reading habits or family resources may influence both.
3. People who exercise more tend to report better sleep; this association alone does not prove that exercise caused the improvement.

F. FIND AND CORRECT THE ERROR

1. Direction alone does not determine strength. Strength depends on how closely the points follow a clear trend.
2. Pearson-type linear correlation can be near zero even when a strong nonlinear pattern exists.

G. PEARSON'S CORRELATION COEFFICIENT, R

1. 0.91: positive; -0.74: negative; 0.08: very weak or negligible linear association.
2. r = -0.88, because |-0.88| = 0.88 is closer to 1 than 0.62. The sign gives direction, not strength.
3. There is a very strong negative linear association: as one variable increases, the other tends to decrease. This does not by itself show causation.

H. R MEASURES LINEAR ASSOCIATION

1. The scatter plot shows a clear curved relationship even though positive and negative slopes balance to give $r \approx 0$.
2. There is little or no linear correlation.

Name: _____ Date: _____

UNIT 5

4.5.7 — Correlation

ANSWER KEY

I. COMPARE EVIDENCE

1. **Set A.**
2. **No. r measures strength of linear association, not importance, practical effect, causation or value in context.**
3. **An outlier can sit far from the main linear trend and therefore change the covariance pattern that r summarizes.**

J. REGRESSION LINE IN CONTEXT

1. **For each additional hour studied, the model predicts an increase of about 4.2 score points.**
2. **$y = 4.2(5) + 38 = 59$, so the predicted score is 59.**
3. **It is the model's predicted score at 0 study hours. Whether that interpretation is meaningful depends on whether 0 hours is sensible and near the observed data.**

K. INTERPOLATION AND EXTRAPOLATION

1. **Interpolation.**
2. **Extrapolation.**
3. **12 hours is well outside the observed range, so the linear pattern may not continue that far.**

L. CHALLENGE — EVALUATE A CLAIM

1. **First, correlation does not prove causation. Second, a regression/correlation statement describes a tendency, not a guaranteed outcome for every individual.**
2. **r only summarizes linear association. A strong U-shaped relationship can therefore produce a small r-value.**
3. **Example checklist: (1) inspect the scatter plot for form, direction, strength and unusual points; (2) interpret r or a regression model only if a linear model is sensible; (3) state conclusions in context without claiming causation and treat extrapolation cautiously.**

TEACHER NOTE

Sections G–L are clearly labelled extensions beyond the supplied 4.5.7 source deck. The source deck itself focuses on qualitative direction/strength, contextual interpretation and the distinction between correlation and causation.