

Algebra 1 • Unit 3 Worksheet

Two-Variable Statistics

Name: _____

Date: _____

Part A • Two-Way Tables

1. A survey of 200 students asked whether they prefer dogs or cats and whether they live in a house or an apartment.

- What fraction of HOUSE dwellers prefer dogs? _____
- What fraction of DOG-LOVERS live in houses? _____
- Why are (a) and (b) DIFFERENT, even though both involve "dogs" and "houses"? Explain in your own words.

d. Is there an association between housing type and pet preference here? Justify using relative frequencies.

	Dog	Cat	Total
House	84	36	120
Apartment	32	48	80
Total	116	84	200

Part B • Scatter Plots and Linear Models

2. A scatter plot shows shoe size (x) vs. height in inches (y) for 12 students. The best-fit line is $y = 1.4x + 50$.

- What does the SLOPE 1.4 mean in context? (Use units.)

b. What does the y-INTERCEPT 50 mean? Is that interpretation sensible here? Explain.

c. Predict the height of a person with shoe size 11. _____ inches

d. Predict the height of a person with shoe size 25. Would you trust this prediction? Why or why not?

Part C • Residuals and Goodness of Fit

3. For a data set, the linear model is $\hat{y} = 2x + 5$.

a. An observation is (3, 13). Compute the residual. Residual = _____

b. Another observation is (7, 17). Compute the residual. Residual = _____

c. What does it mean if a residual is POSITIVE vs. NEGATIVE?

d. If MOST residuals are clustered near zero, what does that suggest about the model?

4. Two scatter plots both have correlation coefficient $r \approx 0.85$, but one has a clearly CURVED pattern and the other looks linear with random scatter.

Which one is better described by a linear model? How can residual analysis help you decide?

★ This is a classic warning: high $|r|$ ALONE doesn't guarantee a good linear fit.

Part D • Correlation Coefficient

5. Match each scatter plot description with a likely correlation coefficient. Choose from: $r \approx -0.95$, $r \approx -0.3$, $r \approx 0$, $r \approx 0.4$, $r \approx 0.97$

- a. Strong, almost-perfect upward trend. $r \approx$ _____
- b. Strong downward trend with small scatter. $r \approx$ _____
- c. No clear pattern; points look random. $r \approx$ _____
- d. Weak positive trend; lots of scatter. $r \approx$ _____
- e. Weak negative trend; lots of scatter. $r \approx$ _____

Part E • Correlation vs. Causation

6. A researcher finds that as ice cream sales rise in a city, drowning deaths also rise. The correlation is strong and positive.

- a. Does ice cream cause drowning? Why might these two be correlated anyway?

-
- b. Name the third variable that could explain BOTH trends.

7. For each pair, decide whether the relationship is most likely CAUSAL, MERELY CORRELATED (common cause), or COINCIDENTAL. Briefly justify.

- a. Hours studied vs. exam score.
-

b. Number of firefighters at a fire vs. damage caused.

c. Number of pirates worldwide vs. global average temperature.

d. Daily calorie intake vs. weight gained.

Part F • Error Analysis and Synthesis

8. Lin says: "The correlation between my friends' shoe size and their math test score is 0.4. This means bigger feet cause better math grades."

a. What is wrong with Lin's claim? Identify at least TWO problems.

b. Suggest a more reasonable interpretation.

9. A study finds $r = 0.99$ between two variables.

a. Should we automatically use a linear model to predict? Why or why not?

b. What other check would you do before trusting the model?

10. Open: Design your own study.

a. Identify two numerical variables you think might be related:

b. What sign of r do you predict? Strong or weak?

c. Could there be a confounding (third) variable? Name one.

d. If you found $r = 0.6$, could you claim one CAUSES the other? Why or why not?
