Using Linear Regression to Explore Environmental Factors Affecting Vector-borne Diseases

ASSESSMENT QUESTIONS

This document contains example assessment questions that you may choose to use to determine whether or not your students have met the Learning Objectives (stated in the Instructor Guide).

1. Describe a situation where another factor could be masking the relationship between two variables.

   Solution: Answers may vary. Example- Lyme Disease cases and drowning deaths →

   Third variable Temperature/Season of the year

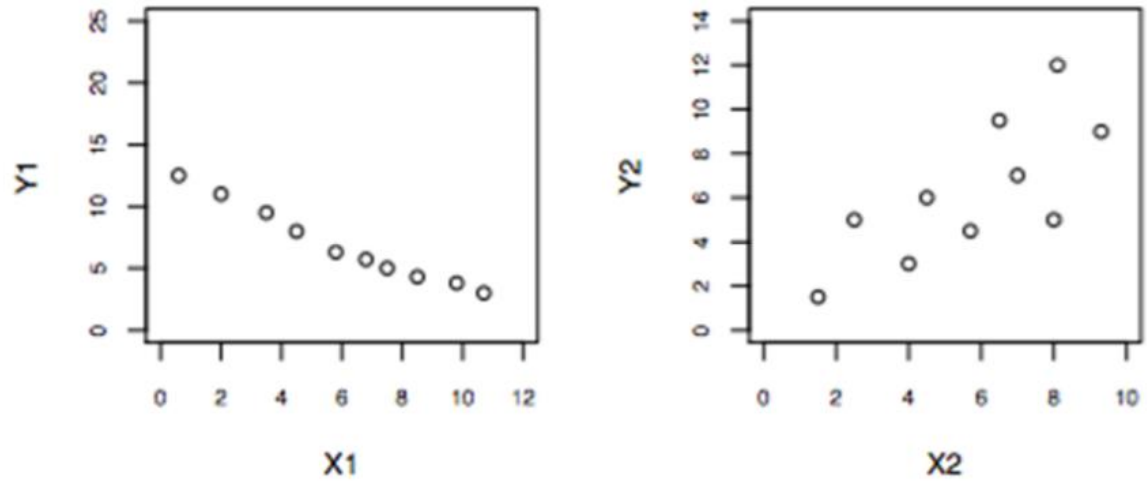
2. A recent article in a Biology research journal reports a correlation of +0.8 between air quality and trees per square mile. It also reports a correlation of -0.8 between trees per square mile and humans per square mile. Which of the following interpretations is the most correct?
   A. The correlation of +0.8 indicates a stronger relationship than the correlation of -0.8.
   B. The correlation of +0.8 is just as strong as the correlation of -0.8.
   C. It is impossible to tell which correlation is stronger.
   Solution: B.

3. A study found that individuals who lived in houses with more than two televisions tended to live longer than individuals who lived in houses with two or fewer televisions. Can a cause-and-effect conclusion be drawn from this? Why or why not?
   Solution: No, this is an observational study and causation can not be proven with an observational study. There is only an association. One reason for this association could be that individuals who can afford multiple televisions most likely have access to health care.

4. The linear model $y = 1.648(x) + 3.571$ represents the number of Bourbon Disease cases (y) given the temperature (x) in degrees F. What is the correct interpretation of the slope of this model?
   a. For every 1.648 cases, the temperature increases 3.571 degrees.
   b. For every one degree increase in temperature, the number of cases increases by 3.571.
   c. For every one degree increase in temperature, the number of cases increases by 1.648.
   d. For every 3.571 cases, the temperature increases 1.648 degrees.

   Solution: C.

5. Give a rough estimate of the sample correlation for the data in each of the scatterplots below.

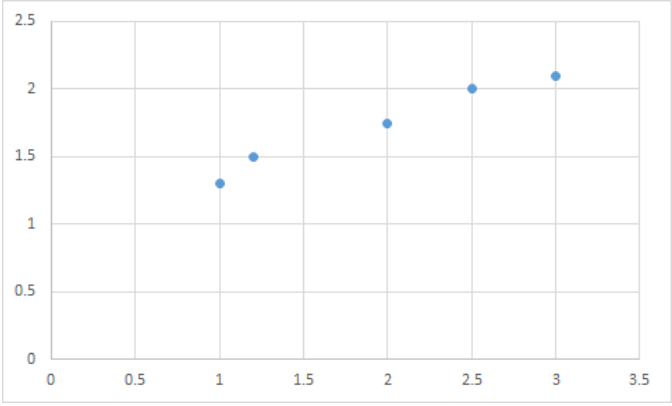


Solution: Answer may vary. Example: Graph on the left: r = -0.9, Graph on right: r = 0.65

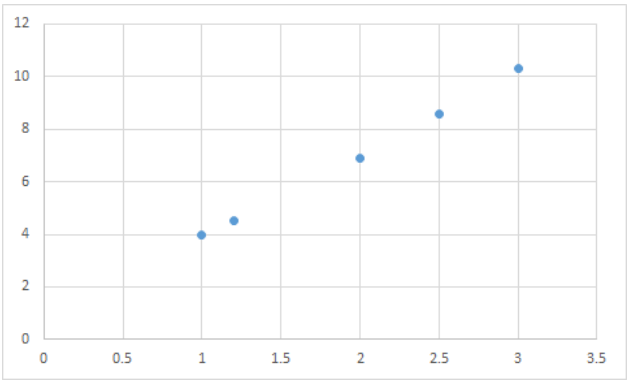
6. For each part, draw a scatterplot satisfying the conditions given, or else explain why the conditions are impossible:
   - Regression line has small positive slope and correlation is high and positive.
   - Regression line has large positive slope and correlation is high and positive.
   - Regression line has small positive slope and correlation is low and positive.
   - Regression line has large positive slope and correlation is low and positive.
   - Regression line has positive slope and correlation is negative.

Solution:

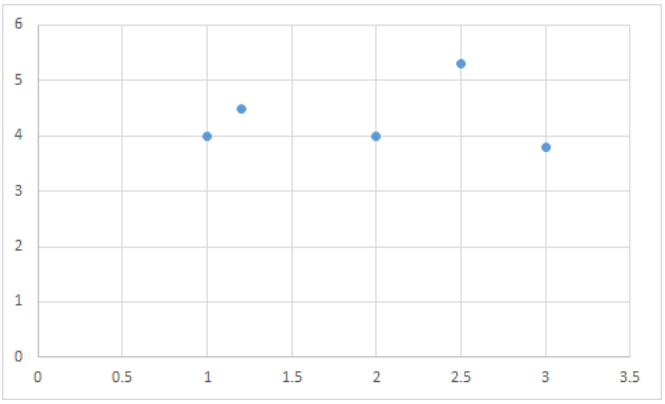
- Regression line has small positive slope and correlation is high and positive.



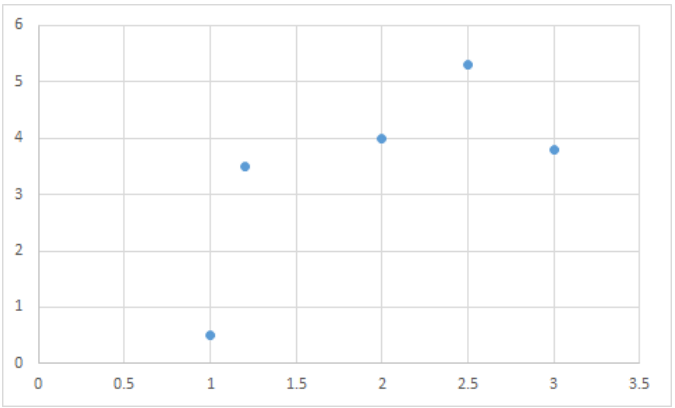
- Regression line has large positive slope and correlation is high and positive.



- Regression line has small positive slope and correlation is low and positive.



- Regression line has large positive slope and correlation is low and positive.



- Regression line has positive slope and correlation is negative. Impossible, a positive slope results in a positive correlation.

7. Researchers collect data to investigate the correlation between water pollution and pond biodiversity. The found a correlation coefficient of r = - 0.68. Select the statement that accurately summarizes this information.
    a. As water pollution increases, pond biodiversity stays the same
    b. As water pollution increases, pond biodiversity decreases
    c. As water pollution increases, pond biodiversity increases

Solution: B The negative correlation coefficient indicates a downward trendline. As the independent variable (water pollution) increases, the dependent variable (pond biodiversity) decreases.

8. Consider a new viral disease that infects elk in the conifer forests of North America. The vector for this disease is an insect that thrives in cool, short summers. Researchers found that temperatures above 75 °F hinder the reproductive success of this insect. Predict the effect of climate change on the spread of this disease.
   Solution: Climate change would cause an increase in temperatures in the conifer forest of North America. This would potentially decrease the population of this vector. Therefore, the disease transmission would likely decrease.

9. The linear model y = 2.6(x) + 8 represents the number of mosquitoes in a certain area and (y) given the temperature (x) in degrees F. What is the correct interpretation of the y-intercept of this model?
   a. When the number of the mosquitoes is 8, the temperature is 2.6.
   b. When the temperature is 0, the number of mosquitoes is 8.
   c. For every one degree increase in temperature, the number of mosquitoes increases by 2.6.
   d. When the number of the mosquitoes is 0, the temperature is 2.6.

   Solution: B

10. Describe a limitation in the data or data analysis for the activity that you completed. As a reminder your activity investigated the relationship between a specific disease and temperature or rainfall.
    Answers May Vary. Possible Answer: The data sets have a small number of points, and data with a large set of points would represent the relationships more accurately. Also there can be errors in the data collection. The relationships considered may not be linear and could show a stronger relationship using log or other transforms of the data.

Co-Authors (Alphabetical): Andy Adams, Jessica Adams, John Bray, Tami Imbierowicz, Suzanne Lenhart, Breonna Martin. All co-authors contributed equally. This material is based upon work supported by the National Science Foundation under Grant No. 1919613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.