Linear Transformation on Bivariate Data

Statistics Worksheet · Grade 11–12

Name: ______________________________ Date: __________________

Learning Objectives

- Explain what it means to transform variables in a bivariate data set
- Distinguish between linear and nonlinear transformations and their effect on the correlation coefficient
- Apply linear and nonlinear transformations to achieve linearity in a bivariate relationship

Problems

1. Which of the following operations represents a LINEAR transformation on a variable? Select all that apply: (A) multiplying the variable by 5, (B) taking the square root of the variable, (C) adding 10 to the variable, (D) raising the variable to the power of 3.

2. A researcher collects bivariate data and calculates a correlation coefficient of r = 0.87. She then multiplies every x-value by 3 and adds 5 to every y-value. What is the new correlation coefficient after this linear transformation?

$$r = 0.87$$

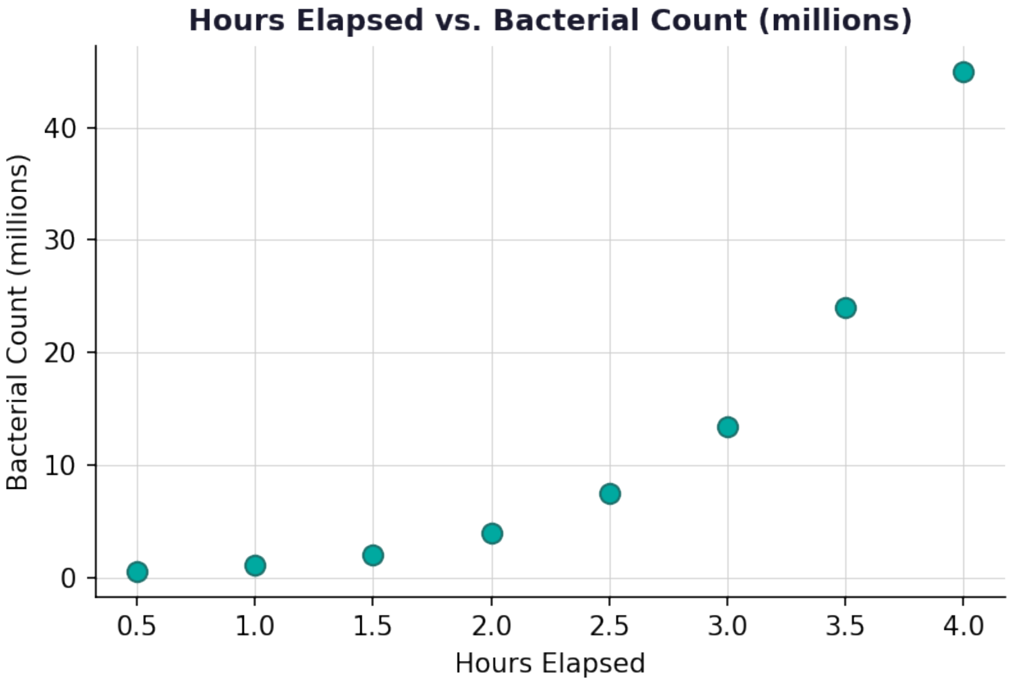
3. A student applies a nonlinear transformation (taking the logarithm of the response variable) to a bivariate data set. Before the transformation, r = 0.62. After the transformation, r = 0.95. Did the transformation improve or weaken the linear relationship? Explain.

4. Given the standard linear regression model below, identify the y-intercept and slope, then predict the response variable when x = 4.

$$\hat{y} = -4.2744 + 3.9433x$$

5. The scatter plot below shows the relationship between hours elapsed (x) and bacterial count in millions (y). Describe the shape of the relationship and explain whether a linear model is appropriate.

Scan to watch



6. The table below shows bacterial count data. Complete the table by calculating log(y) for each y-value, rounding to 4 decimal places. This is the first step in an exponential transformation.

Hours (x)	Bacteria in millions (y)	log(y)
1	2	
2	4	
3	8	
4	16	
5	32	

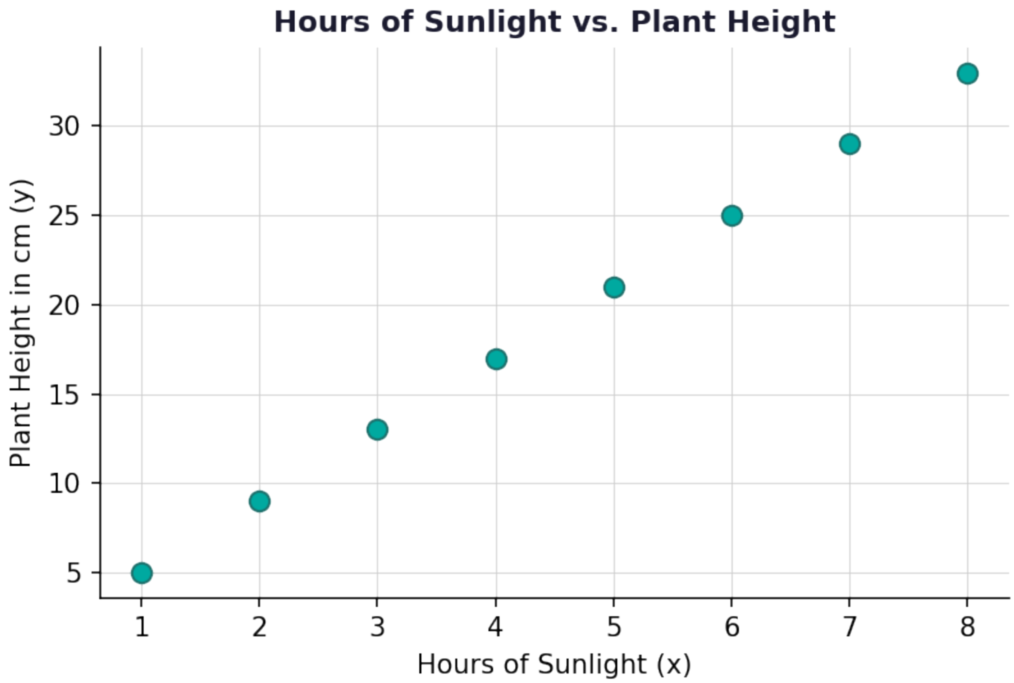
7. After applying a logarithmic transformation to the response variable (bacterial count), the new linear model using log(y) and x is given below. Use this model to predict the bacterial count (in millions) at x = 3.75 hours. Round your answer to 2 decimal places.

$$\log(\hat{y}) = 0.0223 + 0.3010x$$

Scan to watch

8. A scatter plot of x versus y shows a nonlinear (curved) pattern. A student transforms the data by squaring each x-value to get x-squared. After plotting x-squared versus y, the pattern becomes linear with correlation r = 0.98. Explain what type of transformation was applied and whether it was effective. Also state what the original relationship between x and y likely looks like.

9. The scatter plot below shows hours of sunlight (x) versus plant height in centimeters (y) for 8 plants. Find the least squares regression line for this data, interpret the slope in context, and calculate the residual for the point (5, 22).



10. A bivariate data set shows a curved relationship between x (time in years) and y (population in thousands). Two models are proposed: Model A uses linear regression on the original data and gives r = 0.73. Model B applies a log transformation to y, then fits a linear regression to x and log(y), giving r = 0.99. The linear regression equation for Model B (after back-transformation) is given below. Using Model B, predict the population at x = 12 years. Then explain why Model B is preferred over Model A and describe one limitation of extrapolating with Model B.

$$\log(\hat{y}) = 1.20 + 0.085x$$

Scan to watch

Linear Transformation on Bivariate Data — Answer Key

Statistics Worksheet · Grade 11–12

Answer Key

1. Answer: A and C

- Linear transformations include operations such as addition, subtraction, multiplication, and division by a constant.
- Multiplying by 5 (A) and adding 10 (C) are linear operations.
- Taking the square root (B) and raising to a power (D) are nonlinear operations.
- Answer: A and C.

2. Answer: r = 0.87

- Multiplying x-values by a constant and adding a constant to y-values are both linear transformations.
- Linear transformations do NOT change the value of the correlation coefficient r.
- Therefore, the new correlation coefficient remains r = 0.87.

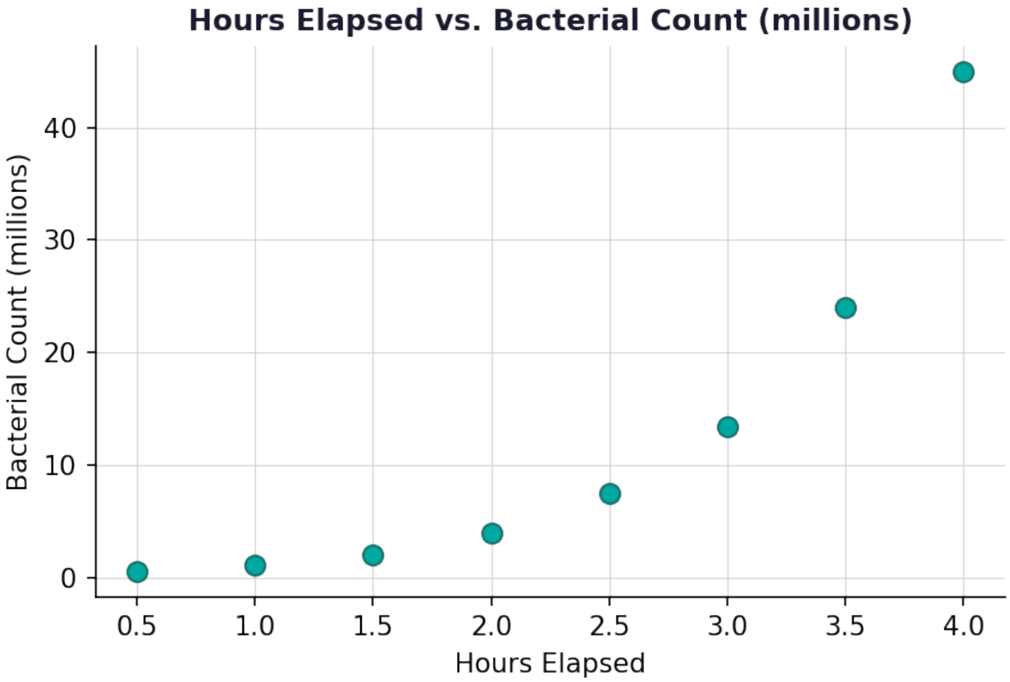
3. Answer: Improved; r increased from 0.62 to 0.95, indicating a stronger linear relationship after transformation.

- Before transformation: r = 0.62 (moderate linear association).
- After nonlinear transformation: r = 0.95 (strong linear association).
- Since |r| increased, the transformation improved linearity.
- This is why nonlinear transformations are used — to achieve a stronger linear relationship.

4. Answer: y-intercept = -4.2744, slope = 3.9433, predicted value at x = 4 is 11.499

- The model is in the form y-hat = a + bx.
- y-intercept (a) = -4.2744.
- Slope (b) = 3.9433.
- Substitute x = 4: y-hat = -4.2744 + 3.9433(4) = -4.2744 + 15.7732 = 11.499.
- The predicted value is approximately 11.499 million bacteria.

5. Answer: The relationship is curved (exponential), so a linear model is NOT appropriate without transformation.



- The scatter plot shows a curved, upward-bending pattern.
- This indicates an exponential (nonlinear) relationship between hours and bacterial count.
- A straight-line (linear) model would not fit this data well.
- A nonlinear transformation (e.g., taking the log of y) is needed to achieve linearity.

---

6. Answer: log(2)=0.3010, log(4)=0.6021, log(8)=0.9031, log(16)=1.2041, log(32)=1.5051

Hours (x)	Bacteria in millions (y)	log(y)
1	2	0.3010
2	4	0.6021
3	8	0.9031
4	16	1.2041
5	32	1.5051

- log(2) = 0.3010
- log(4) = 0.6021
- log(8) = 0.9031
- log(16) = 1.2041
- log(32) = 1.5051
- Notice that log(y) increases linearly as x increases by 1, confirming an exponential model.

---

7. Answer: Approximately 8.93 million bacteria

- Substitute x = 3.75 into the transformed model: log(y-hat) = 0.0223 + 0.3010(3.75).
- log(y-hat) = 0.0223 + 1.1288 = 1.1511.

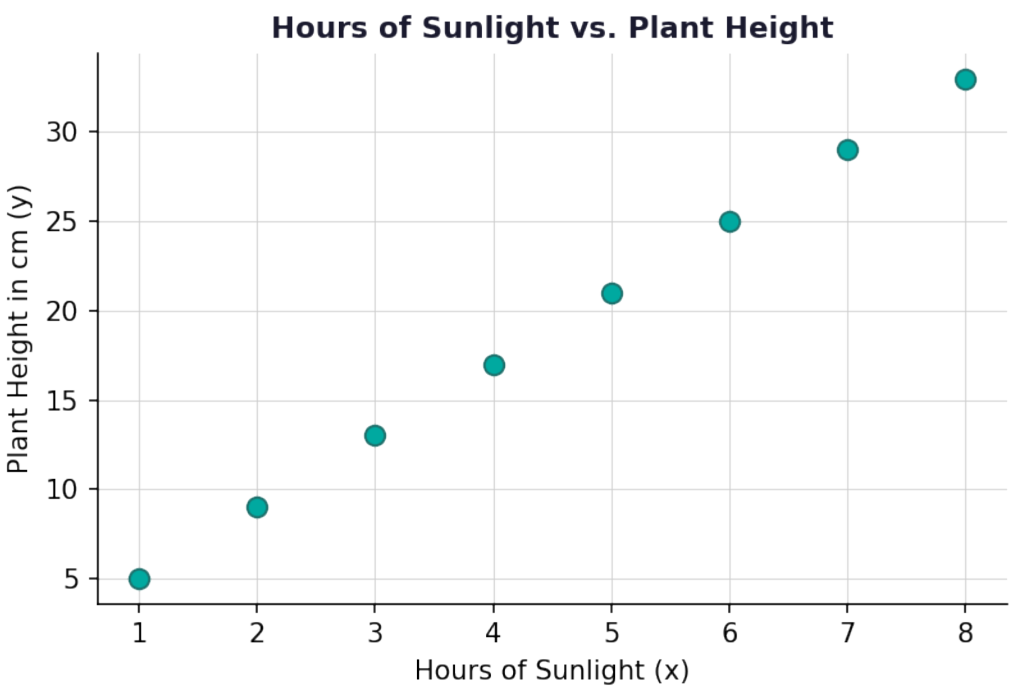
Scan to watch

- To find y-hat, raise 10 to the power of 1.1511: y-hat = 10^(1.1511).
- y-hat ≈ 14.16 million bacteria.
- Note: Verify using your calculator for precision.

8. Answer: A nonlinear (power/quadratic) transformation was applied to x; it was effective (r = 0.98 is very strong). The original relationship is quadratic.

- Squaring x is a nonlinear transformation of the explanatory variable.
- After transformation, plotting x^2 vs. y gives a strong linear correlation (r = 0.98).
- This is effective because the linearity was achieved.
- The original x vs. y relationship is quadratic (parabolic), which is why squaring x linearized it.
- Nonlinear transformations can increase or decrease r depending on the underlying pattern.

9. Answer: Regression line: y-hat = 1 + 4x; slope means each additional hour of sunlight adds 4 cm of height; residual at (5, 22) = 22 - 21 = 1



- The data shows a perfect linear pattern. Using least squares regression: y-hat = 1 + 4x.
- Slope = 4: for each additional hour of sunlight, plant height increases by 4 cm on average.
- y-intercept = 1: predicted height with 0 hours of sunlight is 1 cm.
- At x = 5: y-hat = 1 + 4(5) = 21 cm.
- Residual = observed - predicted = 22 - 21 = 1 cm.
- A positive residual means the actual height was 1 cm more than predicted.

10. Answer: Predicted population ≈ 164.8 thousand; Model B preferred because r = 0.99 >> 0.73; limitation: exponential growth may not hold far beyond data range.

- Substitute x = 12 into Model B: log(y-hat) = 1.20 + 0.085(12) = 1.20 + 1.02 = 2.22.
- Back-transform: y-hat = 10^(2.22) ≈ 165.96 thousand (approximately 166 thousand).

Scan to watch

• Model B is preferred because its correlation after transformation (r = 0.99) is much closer to 1 than Model A's r = 0.73.
• A higher |r| after transformation means the transformed model fits the data far better.
• Limitation of extrapolation: the exponential growth pattern assumed by the log model may not continue far outside the range of observed data, leading to unreliable predictions.