

Key

Ch 4 REVIEW: Linear Regression Name:

Part 1: Multiple-choice. Circle the letter corresponding to the best answer choice.

A study gathers data on the outside temperature during the winter, in degrees Fahrenheit, and the amount of natural gas a household consumes, in cubic feet per day. Call the temperature x and gas consumption y . The house is heated with gas, so x helps explain y . The least-squares regression line is $y = 1.344x - 19$. The next two questions concern this line.

1. On a day when the temperature is 43°F , the regression line predicts that gas used will be about
 $y = 1.344(43) - 19 = 48.9$ ft³

2. When the temperature goes up 1 degree, what happens to the gas usage predicted by the regression line?
 gas use goes up 1.344 ft³

3. The correlation between temperature x and gas usage y is $r = -0.7$. What is the effect of changing Celsius to Fahrenheit on the correlation?
 No effect

4. All 753 students in grades 1 through 6 in an elementary school are given a math test that was designed for third graders. The body weights of all 753 students are also recorded. What kind of association can we expect to see between the weight of a student and their test score?
 NONE

5. A study of the effects of television measured how many hours of television each of 125 grade school children watched per week during a school year and their reading scores. Which variable would you put on the horizontal axis of a scatterplot of the data?
 Hours of TV

6. The study described in the previous question found that children who watch more television tend to have lower reading scores than children who watch fewer hours of television. The study report says that "Hours of television watched explained 15% of the observed variation in the reading scores of the 125 subjects." What is the correlation between hours of TV and reading score?
 $r^2 = 0.15$ $r = \sqrt{0.15} = 0.39$

7. There is a close relationship between the correlation r and the slope b of the least-squares regression line. What is it that they have in common?
 direction (pos or neg)

8. If there were something genetic that made people simultaneously more susceptible to both smoking and lung cancer, this would be an instance of lurking variable.

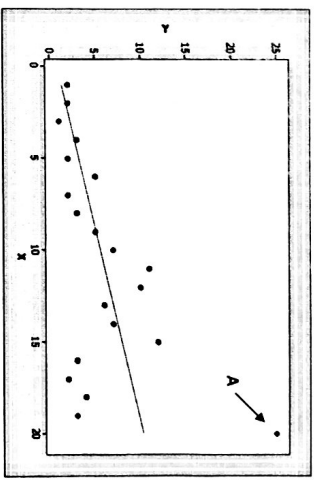
9. The United Nations has data on the percent of adult males and females who are illiterate in each of these 142 countries. The correlation between male illiteracy rate and female illiteracy rate is $r = 0.945$. What does this r value tell us?
 r is positive, there is a pos strong relationship b/w male and female illiteracy.

10. Give an example of strong negative correlation, positive correlation and no correlation.
 -0.9, 0.8, 0

x	23	15	26	24	22	29	32	40	41	46
y	19	18	22	20	27	25	32	38	35	45

11. The regression line for the data given above is $y = 2.35 + 0.86x$. What is the residual for the point whose x -value is 15?
 Predict $y = 2.35 + 0.86(15) = 15.25$ Residual =
 Actual (15, 18) $18 - 15.25 = 2.75$

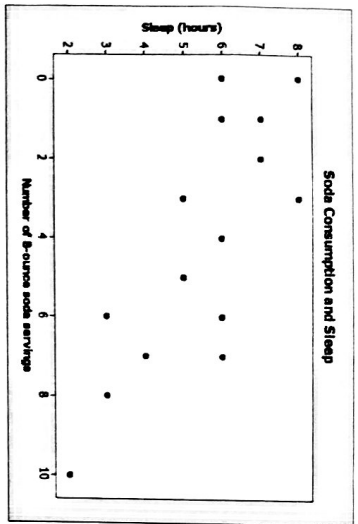
12. A scatterplot and a least-squares regression line are shown in the figure below. If the point (20, 25) that is labeled A is removed from the data set, what happens to the slope, the y -intercept, and the correlation?



Slope \rightarrow smaller
 y -int - No change
 correlation - stronger

13 A random sample of boarding school students was asked how many 8-ounce servings of soda they had consumed on a certain Sunday and how many hours of sleep they got that night. Their responses are displayed in the scatter plot below.

Soda	Sleep
0	6
0	8
1	6
1	7
2	7
3	5
3	8
4	6
5	5
6	3
6	6
7	4
7	6
8	3
10	2



(a) Which variable has been used as the explanatory variable in making the scatterplot above? Could the two variables have been reversed? Explain.

Exp = # sodas
 No → # sodas won't be changed by sleep tonight

(b) Describe what the scatterplot tells you about the direction, form, and strength of the relationship between the variables.

Negative, linear, moderate

(c) What is the slope and y-intercept of the least-squares regression line for this data? Interpret these values in the context of the problem.

$$y = -0.416x + 7.39$$

Slope = -0.416
 For each soda, hours of sleep decreases by 0.416.

(d) Suppose that a student had 10 servings of soda that Sunday. What would you predict for the number of hours the student slept that night? Show your work.

$$y = -0.416(10) + 7.39 = 2.79 \text{ hours}$$

For 0 sodas, hours of sleep = 7.39.

(e) Interpret the meaning of r and r^2 in this setting.

$r = -0.78$ neg, moderate relationship b/w sodas and sleep

$r^2 = 0.61$ 61% of change in hours of sleep can be explained by # of sodas.