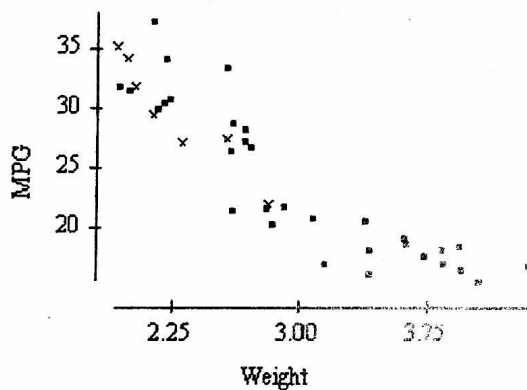


## Chapter 3: Examining Relationships **Review**

1. A study is conducted to determine if one can predict the yield of a crop based on the amount of yearly rainfall. The response variable in this study is
  - A) yield of the crop.
  - B) amount of yearly rainfall.
  - C) the experimenter.
  - D) either bushels or inches of water.
2. The graph below plots the gas mileage (miles per gallon, or MPG) of various 1978 model cars versus the weight of these cars in thousands of pounds.



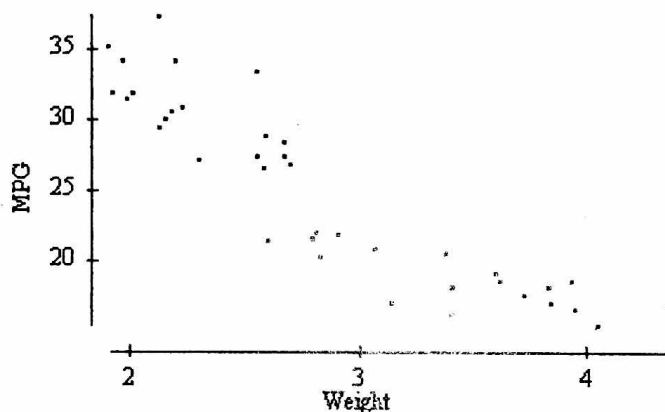
In the graph, the points denoted by the plotting symbol x correspond to cars made in Japan. From this plot, we may conclude that

- A) in 1978 there was little difference between Japanese cars and cars made in other countries.
- B) in 1978 Japanese cars tended to be lighter in weight than other cars.
- C) in 1978 Japanese cars tended to get poorer gas mileage than other cars.
- D) the plot is invalid. A scatterplot is used to represent quantitative variables, and the country that makes a car is a qualitative variable.

3. A school guidance counselor examines the number of extracurricular activities of students and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero." A correct interpretation of this statement would be that

- A) active students tend to be students with poor grades, and vice versa.
- B) students with good grades tend to be students that are not involved in many activities, and vice versa
- C) students involved in many extracurricular activities are just as likely to get good grades as bad grades. The same is true for students involved in few extracurricular activities.
- D) involvement in many extracurricular activities and good grades go hand in hand.

4. Consider the scatterplot below.



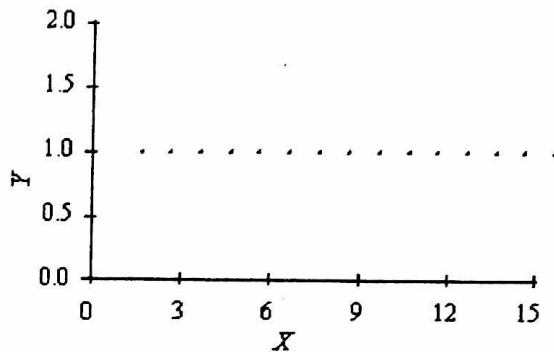
According to the scatterplot, which of the following is a plausible value for the correlation coefficient between weight and MPG?

- A) +0.2.
- B) -0.9.
- C) +0.7.
- D) -1.0.

5. Which of the following is true of the correlation coefficient  $r$ ?

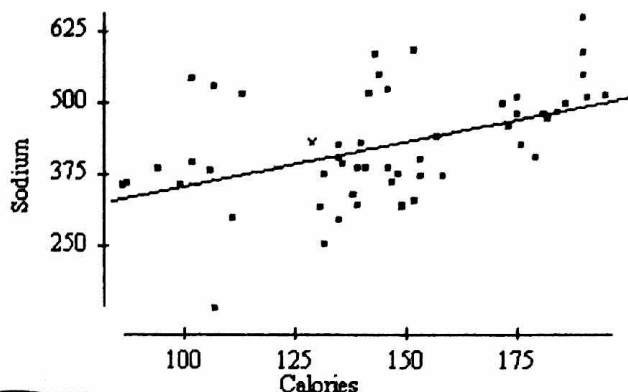
- A) It is a resistant measure of association.
- B)  $-1 \leq r \leq 1$ .
- C) If  $r$  is the correlation between  $X$  and  $Y$ , then  $-r$  is the correlation between  $Y$  and  $X$ .
- D) All of the above.

6. A scatterplot of a variable  $Y$  versus a variable  $X$  produced the scatterplot below. The value of  $Y$  for all values of  $X$  is exactly 1.0. The correlation between  $Y$  and  $X$  is



- A) +1.0, because the points lie perfectly on a line.  
B) either +1.0 or -1.0, because the points lie perfectly on a line.  
C) 0, because  $Y$  does not change as  $X$  increases.  
D) none of the above.
7. The fraction of the variation in the values of  $y$  that is explained by the least-squares regression of  $y$  on  $x$  is
- A) the correlation coefficient.  
B) the slope of the least-squares regression line.  
C) the square of the correlation coefficient.  
D) the intercept of the least-squares regression line.
8. The least-squares regression line is
- A) the line that makes the square of the correlation in the data as large as possible.  
B) the line that makes the sum of the squares of the vertical distances of the data points from the line as small as possible.  
C) the line that best splits the data in half, with half of the points above the line and half below the line.  
D) all of the above.

9. A scatterplot of the calories and sodium content of several brands of meat hot dogs is shown below. The least-squares regression line has been drawn in on the plot. Referring to this scatterplot, the value of the residual for the point labeled x



- A) is about 40.  
 B) is about 1300.  
 C) is about 425.  
 D) cannot be determined from the information given.

*Actual - Predicted*

10. A researcher wishes to determine whether the rate of water flow (in liters per second) over an experimental soil bed can be used to predict the amount of soil washed away (in kilograms). The researcher measures the amount of soil washed away for various flow rates and from these data calculates the least-squares regression line to be
- $$\text{amount of eroded soil} = 0.4 + 1.3 \times (\text{flow rate})$$
- One of the flow rates used by the researcher was 0.3 liters per second; for this flow rate the amount of eroded soil was 0.8 kilograms. These values were used in the calculation of the least-squares regression line. The residual corresponding to these values is

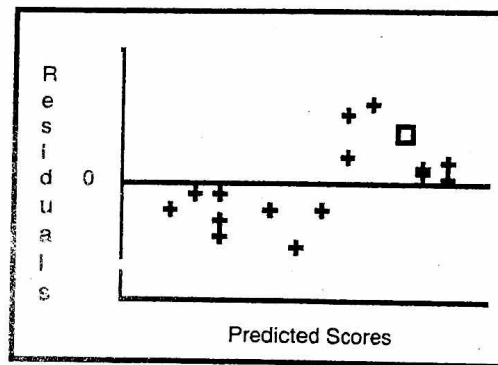
- A) 0.01  
 B) -0.01.  
 C) 0.5.  
 D) -0.5.

## Free Response

- Given a bivariate dataset such that  $\bar{x} = 14.5$ ,  $\bar{y} = 20$ ,  $s_x = 4$ ,  $s_y = 11$ ,  $r = .80$ . Find the least-squares regression line of  $y$  on  $x$ .
- The data given below give the first and second exam scores of 10 students in a calculus class

Test 1	63	32	87	73	60	63	83	80	98	85
Test 2	51	21	52	90	83	54	73	85	83	46

- Draw a scatterplot of these data.
  - To what extent do the scores on the two tests seem related?
- The following is a residual plot of a linear regression. Clearly, a line would not be a good fit for these data. Why not? Is the regression equation likely to underestimate or overestimate the  $y$ -value of the point in the graph marked with the square?



- The regional champion in 10 and under 100 m backstroke has had the following winning times (in seconds) over the past 8 years:

Year	1	2	3	4	5	6	7	8
Time	77.3	80.2	77.1	76.4	75.5	75.9	75.1	74.3

How many years until you expect the winning time to be one minute or less? What's wrong with this estimate?

- Measurements are made of the number of cockroaches present, on average, every 3 days, beginning on the second day, after apartments in one part of town are vacated. The data are as follows:

Days	2	5	8	11	14
# Roaches	3	4.5	6	7.9	11.5

How many cockroaches would you expect to be present after 9 days?

6. A study found a strongly positive relationship between number of miles walked per week and overall health. A local news commentator, after reporting on the results of the study, advised everyone to walk more during the coming year because walking more results in better health. Comment on the reporter's advice.
7. Carla, a young sociologist, is excitedly reporting on the results of her first professional study. The finding she is reporting is that 72% of the variation in math grades for girls can be explained by the girls' socioeconomic status. What does this mean, and is it indicative of a strong linear relationship between math grades and socioeconomic status for girls?
8. Which of the following statements are true of a least-squares regression equation?
- It minimizes the sum of the residuals.
  - The average residual is 0.
  - It minimizes the sum of the squared residuals.
  - The slope of the regression line is a constant multiple of the correlation coefficient.
  - The slope of the regression line tells you how much the response variable will change for each unit change in the explanatory variable.
9. Consider the following dataset:

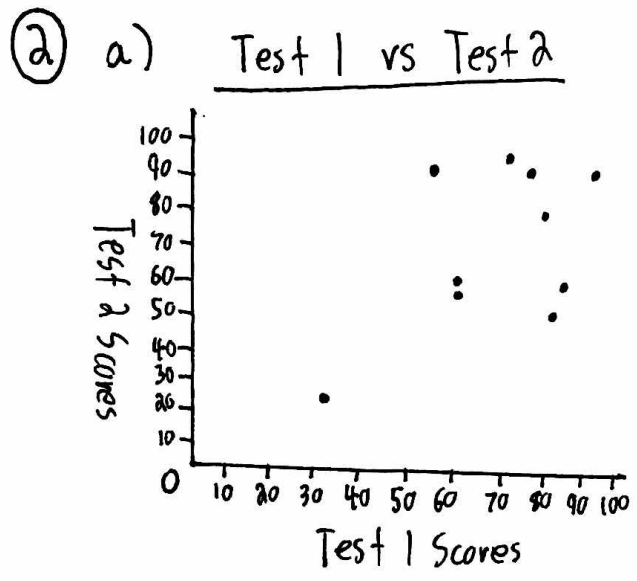
$x$	45	73	82	91
$y$	15	7.9	5.8	3.5

Given that the LSRL for this data is  $\hat{y} = 26.211 - 0.25x$ , what is the value of the residual for  $x = 73$ ?

10. Suppose the correlation between two variables is  $r = -0.75$ . What is true of the correlation coefficient and the slope of the regression line if
- each of the  $y$  values is multiplied by  $-1$ .
  - the  $x$  and  $y$  variables are reversed.
  - the  $x$  and  $y$  variables are each multiplied by  $-1$ .
11. Suppose the regression equation for predicting success on a dexterity task ( $y$ ) from number of training sessions ( $x$ ) is  $\hat{y} = 45 + 2.7x$  and that  $\frac{s_y}{s_x} = 3.33$ .

What percentage of the variation in  $y$  is not explained by the regression on  $x$ ?

①  $\hat{y} = a + bx$  |  $a = 20 - 2.2(14.5) = -11.9$  |  $\hat{y} = -11.9 + 2.2x$   
 $\hat{y} = b_0 + b_1x$  |  $b = 0.80(1/4) = 2.2$



b) The relationship between the first and second exam scores appears to be moderate, positive, and linear. Students who did better on the first test tend to do better on the second test. However, the relationship is not very strong ( $r = 0.5493$ ).

③ There is a pattern to the residual plot which indicates that a linear regression is not a good fit. The regression line is underestimating the y-values (actual - predicted).

④ Year 32 it will be less than 60 seconds (calculator). This is an example of extrapolation (going way beyond the domain of the data).

⑤  $\hat{y} = 1.14 + 0.68x \Rightarrow 7.26$  or 7 cockroaches.

⑥ Just because the study found a strong positive relationship between miles walked and overall health, does not mean one causes the other. (The correlation is not causation).

⑦ Response variable = math grades, Explanatory variable = eco. status for girls.  
 $r^2 = 0.72 \Rightarrow r = \pm 0.8485$   
 The r-value indicates a strong linear association between the two variables.

⑧ B, C, D, and E are true (A - minimize as small as possible, not necessarily negative).

⑨  $\hat{y} = 26.211 - 0.25x \Rightarrow 7.9 - 7.961 = -0.061$

⑩ a)  $r = +0.75$ , opposite of original; b)  $r = -0.75$ , slope negative; c)  $r = -0.75$ , same as original.

⑪  $b = r \frac{S_y}{S_x} \Rightarrow r = b \frac{S_x}{S_y}$  |  $r = (2.7)(0.30) = 0.81$  |  $r^2 = 0.66$  | 34% of the variation in y is not explained by x.