Chapter 3: Describing Relationships

1. The following scatterplot describe the relationship between height (in cm) and foot length (also in cm) for 12 randomly selected students from the British Census @ Schools database. Which of the following is the best description of this relationship?



A. There is some random variation, but otherwise height is directly proportional to foot length.

B. There is a roughly linear relationship between height and foot length.

C. There is a moderately weak, positive linear relationship between height and foot length.

2. A sociologist is studying the relationship between early childhood nutrition and academic achievement in middle school among children in a certain city. Which of the following statements about the variable “early childhood nutrition” is correct?

A. Early childhood nutrition is a response variable.

B. Early childhood nutrition is an explanatory variable.

C. Since there is not a clear explanatory-response relationship in this scenario, we cannot classify early childhood nutrition as either explanatory or response.

3. A sociologist studying the relationship between early childhood nutrition and academic achievement in middle school among children in a certain city finds that the correlation between these two variables is 0.86. Which of the following conclusions can he draw from this study?

A. Ensuring good nutrition in early childhood will increase academic achievement for middle school students.

B. Children in this city who have a healthy diet in early childhood tend to do better in middle school.

C. Since the correlation is so low, no conclusions can be drawn.

4. Which of the following quantities is minimized by the least-squares regression line?

A. The sum of the squared differences between observed values of the response variable and values of the response variable predicted by the model.

B. The sum of the squares of perpendicular distances between all data points and the regression line.

C. The sum of the squared differences between observed values of the explanatory variable and values of the explanatory variable predicted by the model.

5. Which of the following statements about the slope of the least-squares regression line is true?

A. It has the same sign as the correlation coefficient *r*.

B. The square of the slope equals the proportion of the variation in the response variable that is explained by the explanatory variable.

C. It is unitless.

6. The points in the scatterplot represent paired observations (*x*, *y*) where *x* is an individual’s weight and *y* is the time (in seconds) it takes for walking on a treadmill to raise the individual’s pulse rate to 140 beats per minute. The open circles correspond to females and the dark squares to males.



From the scatterplot, which conclusion we can make?

A. There is a positive correlation *r* between gender and weight, since men tend to weigh more than women.

B. There is a negative correlation *r* between weight and time for both males and females.

C. In general, males tend to take less time to have their pulse rate raised to 140 bpm while walking on the treadmill.

7. One of the following is a correct statement involving correlation. The other two contain blunders. Which one is correct?

A. There is a correlation of *r* = 0.54 between the position a football player plays and his or her weight.

B. The correlation between amount of fertilizer and yield of tomatoes was found to be

*r* = 0.33.

C. The correlation between the gas mileage of a car and its weight is *r* = -0.71 gallon-pounds.

8. A study showed that students who spend more time studying for statistics tests tend to achieve better scores on their tests. In fact, the number of hours studied turned out to explain 81% of the observed variation in test scores among the students who participated in the study. What is the value of the correlation between number of hours studied and test score?

A. *r* = 0.81

B. *r* = 0.656

C. *r* = 0.9

9. The following computer output describes the relationship between *y* = height (in cm) and *x*= foot length (also in cm) for 12 randomly selected students from the British Census @ Schools database. The scatterplot for this relationship show a roughly linear shape.

Predictor Coef SE Coef T P

Constant 117.99 28.39 4.16 0.002

Foot length (cm) 1.878 1.155 1.63 0.135

S = 7.39858 R-Sq = 20.9% R-Sq(adj) = 13.0%

Which of the following is an equation of least-squares regression line for these data?

A. 

B. 

C. 

10. The following computer output describes the relationship between *y* = height (in cm) and *x*= foot length (also in cm) for 12 randomly selected students from the British Census @ Schools database. The scatterplot for this relationship show a roughly linear shape.

Predictor Coef SE Coef T P

Constant 117.99 28.39 4.16 0.002

Foot length (cm) 1.878 1.155 1.63 0.135

S = 7.39858 R-Sq = 20.9% R-Sq(adj) = 13.0%

Which of the following is the correct interpretation of the number *s* = 7.39858?

A. For each one-centimeter increase in foot length the model predicts an increase in height of 7.39858 centimeters.

B. If we use the regression equation to predict height from foot length, our predictions will be, on average, off by 7.39858 centimeters.

C. The typical difference between the height of a student and the mean student height is 7.39858.

11. Below is a residual plot for the regression of the number of employees of Microsoft Inc. on year for the years from 1976 to 1989. (Note that this is a residual plot, not a scatterplot!)



There is no observed value for the year 1983. If we were to use this regression to predicted the number of employees in 1983, which of the following is most likely to describe the accuracy of our prediction?

A. Too high

B. Too low

C. About right

12. Ms. Kreppel is interested in the relationship between her students’ final exam scores and their scores on a pre-test they took at the beginning of the year. Below is a scatterplot showing this relationship for the 18 students in her class.



How would the slope of the least-squares regression line change if the individual whose point is circled were removed from the data set?

A. The slope would increase.

B. The slope would decrease.

C. The slope would be unchanged

13. Ms. Kreppel is interested in the relationship between her students’ final exam scores and their scores on a pre-test they took at the beginning of the year. A scatterplot of the data for the 18 students in her class shows linear relationship for these variables. The equation of the least-squares regression line is



Which of the following is a correct interpretation of the slope of this regression model?

A. For each one-unit increase in final exam score, the model predicts, on average, a 0.60 unit increase in pre-test score.

B. For each one-unit increase in pre-test score, the model predicts, on average, a 0.60 unit increase in exam score.

C. About 60% of the variation in exam score that is accounted for by the regression of exam score on pre-test score.

14. Ms. Kreppel is interested in the relationship between her students’ final exam scores and their scores on a pre-test they took at the beginning of the year. A scatterplot of the data for the 18 students in her class shows linear relationship for these variables. The equation of the least-squares regression line is . One student scored a 76 on the pre-test and an 82 on the final exam.

Which of the following is that student’s residual?

A. 2.2

B. –7.4

C. –2.2

15. A residual plot displays a “reverse fan” arrangement, with the spread of points about the line (residual = 0) gradually decreasing from left to right (that is, as *x* increases). Which statement would be a correct interpretation of this plot?

A. The original data display a nonlinear relationship (curved pattern of association).

B. Predictions using the regression line will be more reliable for small *x* than for large *x*.
C. Predictions using the regression line will be more reliable for large *x* than for small *x*.