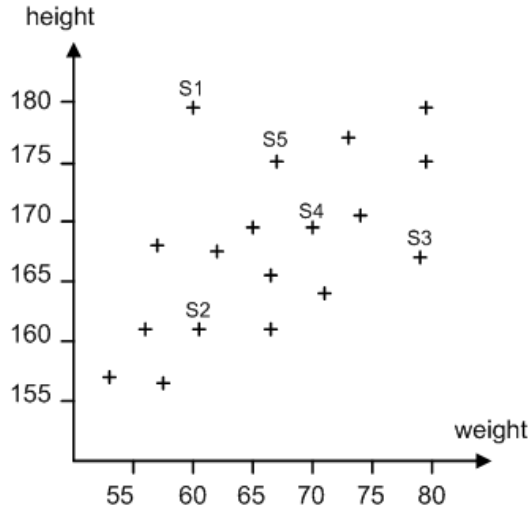
1) A residual plot:

A) displays residuals of the response variable versus the independent variable.
B) displays residuals of the independent variable versus the response variable.
C) displays residuals of the independent variable versus residuals of the response variable.
D) display the independent variable versus the response variable.
E) displays the response variable versus the dependent variable.

2) The height and the weight of 18 students were measured and a scatterplot of the measures is shown below. If two pairs of measurements needs to be removed from the set of 18, which of the choices shown below decreases the coefficient of correlation the most?



A) S_2 and S_3
B) S_2 and S_5
C) S_1 and S_3
D) S_1 and S_2
E) S_2 and S_4

3) The residual value of $(\bar{x}, \bar{y})$ in a linear regression is

A) negative
B) 0
C) positive
D) dependent on the value of r
E) the value cannot be determined

4) If (12, 60) is an influential point for the regression line $\hat{y} = 7.908 + 4.098x$, then which of the following must be true?

A) removal of (12, 60) will improve r
B) removal of (12, 60) will not affect r
C) removal of (12, 60) will change the value of the slope of the regression line
D) (12, 60) has a large residual
E) None of these.

5) Suppose a data set is transformed using (x, y) -> (x, log y) and a least squares linear regression procedure is performed on the transformed data. If the residual plot of this regression shows a curved pattern, which of the following is an appropriate conclusion?

A) A quadratic model should be used with the original data.
B) A square root transformation should be applied to the transformed data.
C) The correlation coefficient of the set of transformed data is 0.
D) The exponential transformation is not appropriate.
E) None of these is appropriate.

6) After data are collected from an agricultural experiment, suppose a transformation is performed on the bivariate set (inches of water, total plant growth). If the linear regression of the transformed data has the equation:

$$\log(\text{growth}) = 0.7 + 1.93\log(\text{water})$$

The regression model of the original data is:

A) growth = 0.7 + 1.93(water)
B) growth = 5.01 + 1.93(water)
C) growth = $(5.01)(1.93)^{\text{water}}$
D) growth = $5.01(\text{water})^{1.93}$
E) None of these

7) Data are obtained from a random sample of adult women with regard to their age and their monthly expenditures on health products. The resulting regression equation is:

$$\text{Expenditures} = 43 + 0.23(\text{Age})$$

with r = 0.27. What percentage of the variation in expenditures can be explained by looking at ages?

A) 0.23%
B) 23%
C) 7.29%
D) 27%
E) 52.0%

8) Data on ages (in years) and prices (in $100) for ten cars of a specific model result in the regression line:

$$\text{Price} = 250 - 30(\text{Age})$$

Given that 64% of the variation in price is explainable by variation in age, what is the value of the correlation coefficient r?

A) -0.64
B) -0.80
C) 0.64
D) 0.80
E) There is insufficient information to answer this question.

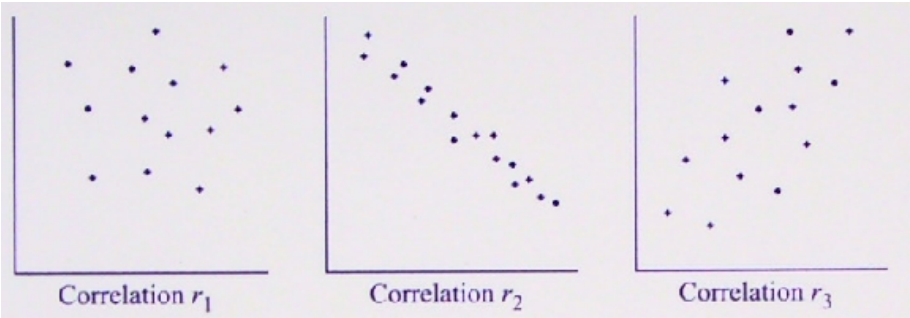
9) Which of the following are true statements about the correlation coefficient r?

I. It is not affected by changes in the measurement units of variables.
II. It is not affected by which variable is called x and which is called y.
III. It is not affected by extreme values.

A) I and II
B) I and III
C) II and III
D) I, II, and III
E) None of the above gives the complete set of true responses.

10) Suppose the correlation between two variables is 0.85. If each of the y-values is multiplied by -1, which of the following is true about the new scatterplot?

A) It slopes up to the right, and the correlation is -0.85.
B) It slopes down to the right, and the correlation is -0.85.
C) It slopes up to the right, and the correlation is 0.85.
D) It slopes down to the right, and the correlation is 0.85.
E) None of the above is true.



11) Which of the following gives a proper ordering?

A) $r_1 < r_3 < r_2$
B) $r_2 < r_3 < r_1$
C) $r_2 < r_1 < r_3 < |r_2|$
D) $r_3 < r_1 < |r_2|$
E) $r_1 < r_2 < r_3 < |r_2|$

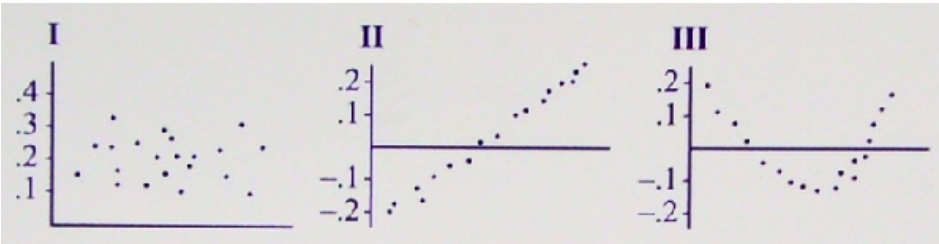
12) A simple random sample of 25 world-ranked tennis players provides the following statistics: Number of hours practiced per day: $\overline{x} = 7.3$, $s_x = 1.2$. Yearly winnings: $\overline{y} = \$1{,}820{,}000$, $s_y = \$310{,}000$. Correlation $r = 0.23$. Based on this data, what is the resulting linear regression equation?

A) Winnings = 1,390,000 + 59,400 hours
B) Winnings = 1,300,000 + 71,300 hours
C) Winnings = -63,400 + 258,000 hours
D) Winnings = -443,000 + 310,000 hours
E) Winnings = -10,000,000 + 1,620,000 hours

13) A linear regression analysis is performed on the data from two scatterplots, A and B, resulting in identical least squares regression lines with positive slopes. Which of the following statements are true?

A) The sum of the squares of the residuals in A equals the sum of the squares in the residuals in B.
B) The correlation in A equals the correlation in B.
C) If the sum of the squares of the residuals in A is greater than the sum of the squares of the residuals in B, then the correlation in A will be greater than the correlation in B.
D) If the sum of the squares of the residuals in A is greater than the sum of the squares of the residuals in B, then the correlation in A will be less than the correlation in B.
E) None of the above are true statements.

14) Which of the following are possible residual plots?



A) I only
B) II only
C) III only
D) I and II
E) I, II, and III

15) Suppose the regression line for a set of data, $\hat{y} = a + 4x$, passes through the point (1, 3). If $\bar{x}$ and $\bar{y}$ are the sample means of the x and y values, respectively, then $\bar{y}$ =

A) $\bar{x}$
B) $4\bar{x}$
C) $3 + 4\bar{x}$
D) $2 + \bar{x}$
E) $-1 + 4\bar{x}$

16) Consider the three points (4, 33), (5, 27), and (6, 15). Given any straight line, we can calculate the sum of the squares of the three vertical distances from these points to the line. What is the smallest possible value this sum can be?

A) 2.45
B) 6
C) 8.66
D) 36
E) None of these values

17) Suppose the correlation between two variables is r = 0.28. What will the new correlation be if 0.17 is added to all the values of the x-variable, every value of the y-variable is doubled, and the two variables are interchanged?

A) 0.28
B) 0.45
C) 0.56
D) 0.90
E) -0.28

18) To the right is a scatterplot with one point labeled X. Suppose you find the least squares regression line. Which of the following statements are true?

I. X has the largest residual, in absolute value, of any point on the scatterplot.
II. X is an influential point.
III. The residual plot will show a curved pattern.

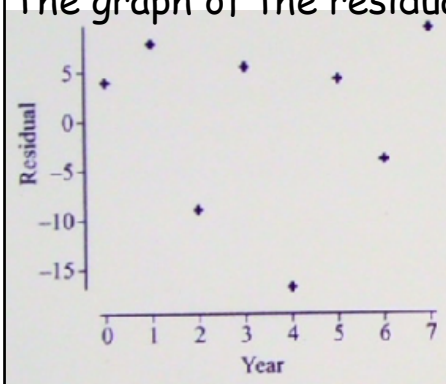
A) I only
B) II only
C) III only
D) I and III
E) II and III

19) Which of the following statements about residuals are true?

I. The mean of the residuals is always zero.
II. The regression line for a residual plot is a horizontal line.
III. The standard deviation of the residuals gives a measure of how the points in the scatterplot are spread around the regression line.

A) I and II
B) I and III
C) II and III
D) I, II, and III
E) None of the above gives the complete set of true responses.

20) The number of students taking AP Stats at a high school during the years 2000-2007 is fitted with a least squares regression line. The graph of the residuals and some computer output is as follows.



Dependent variable is: **Students**
s = 9.758 R-sq = 93.4% R-sq(adj) = 92.4%

Variable	Coeff	s.e.	t	p
Constant	11	6.299	1.75	0.1313
Years	13.9286	1.506	9.25	0.0001

How many students took AP Stats in 2003?

A) 47
B) 48
C) 52
D) 53
E) 58

21) A study of weekly hours of TV watched and SAT scores reports a correlation of r = -1.18. From this information, we can conclude that

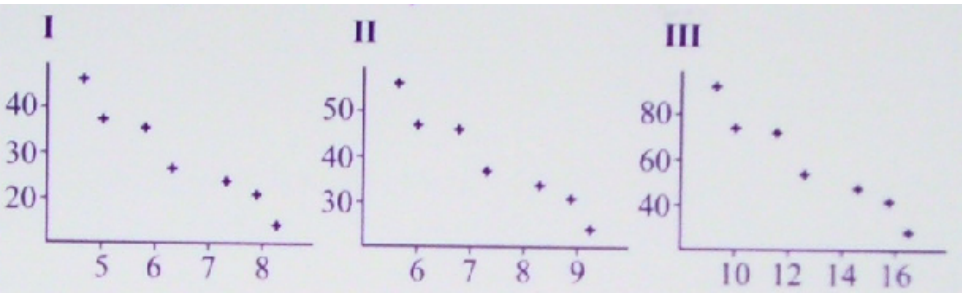
A) students who watch more TV tend to have lower SAT scores.
B) the fewer hours in front of a TV, the higher a student's SAT scores.
C) there is little relationship between weekly hours of TV watched and SAT scores.
D) there is strong negative association between weekly hours of TV watched and SAT scores, but it would be wrong to conclude causation.
E) a mistake in arithmetic has been made.

22) Suppose the scatterplot of log X and log Y shows a strong positive correlation close to 1. Which of the following are true?

I. The variables X and Y will also have a correlation close to 1.
II. A scatterplot of the variables X and Y will show a strong nonlinear pattern.
III. The residual plots of the variables X and Y will show a random pattern.

A) I only
B) II only
C) III only
D) I and II
E) I, II, and III

23) Consider the following three scatterplots:



Which has the greatest correlation coefficient r?

A) I
B) II
C) III
D) The all have the same correlation coefficient.
E) The question cannot be answered without additional information.

24) Which of the following statements about influential points are true?

I. Looking at a residual plot is an excellent way of picking out influential points.
II. Removal of an influential point sharply affects the regression line.
III. Determining a regression model with and without a point is an excellent way of picking out influential points.

A) I and II
B) I and III
C) II and III
D) I, II, and III
E) None of the above gives the complete set of true responses.

25) Which of the following statements about the correlation r is true?

I. The correlation and the slope of the regression line always have the same sign.

II. A correlation of -.32 and of +.32 show the same degree of clustering around the regression line.

III. A correlation of .78 indicates a relationship that is 3 times as linear as one for which the correlation is .26.

A) I and II
B) I and III
C) II and III
D) I, II, and III
E) None of the above gives the complete set of true responses.

Solutions to Chapter 3 "Bluff" Game

Key:
1) A
2) E
3) B
4) C
5) D
6) D
7) C
8) B
9) A
10) B
11) C
12) A
13) E
14) C
15) E
16) B
17) A
18) C
19) D
20) E
21) E
22) B
23) D
24) C
25) A