Name: _____ Class: _____ Date: _____

Chapter 2

1. When creating a scatterplot, one should use the _____ axis for the explanatory variable if a regression line is to be fit to the data.
ANSWER: horizontal

2. A study is conducted to determine whether one can predict the yield of a crop based on the amount of yearly rainfall. The variable _____ is the response variable in this study.
ANSWER: yield of the crop

3. A researcher is interested in determining whether one can predict the score on a statistics exam from the amount of time spent studying for the exam. The variable _____ is the explanatory variable in this study.
ANSWER: amount of time spent studying for the exam

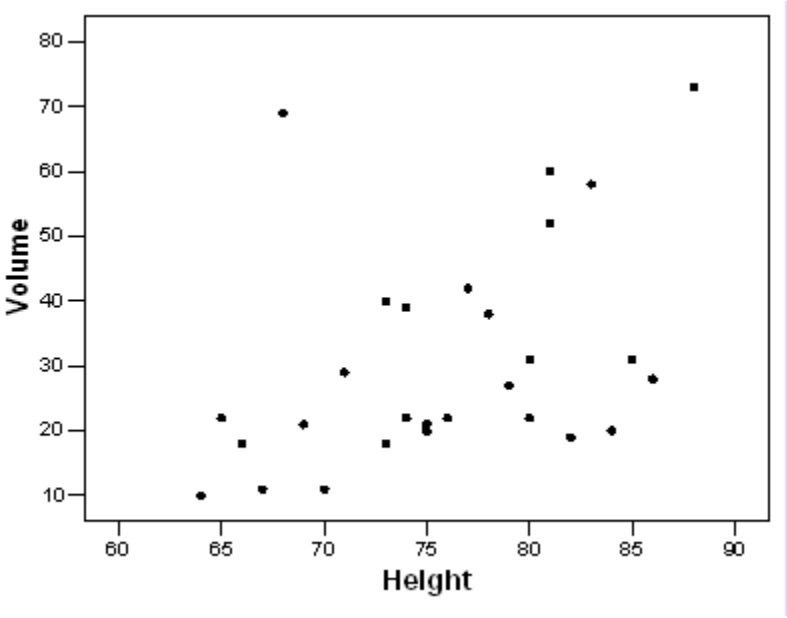
4. The Environmental Protection Agency records data on the fuel economy of many different makes of cars. It is interested in determining whether one can predict the mileage of the car (in miles per gallon) from the weight of the car (in pounds). The variable _____ is the response variable in this study.
ANSWER: mileage of the car

5. The owner of a winery collects data on competing wineries every year. He would like to predict the gross sales (in number of cases) from the size of the winery (in acres). The variable _____ is the explanatory variable in this study.
ANSWER: size of the winery

6. A scatterplot is a graphical tool for displaying the relationship between two _____ variables measured on the same individuals.
ANSWER: quantitative

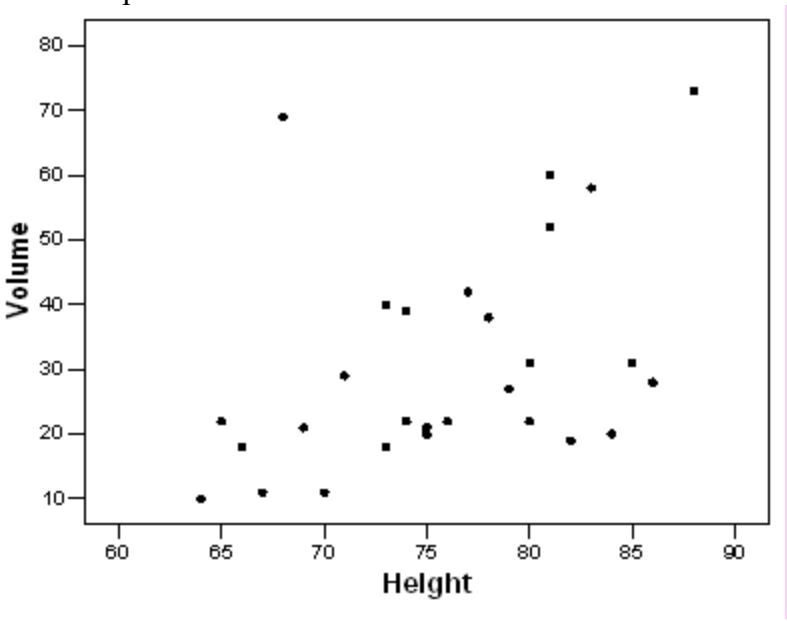
7. A researcher measured the height (in feet) and the volume (in cubic feet) of usable lumber of 32 cherry trees. The goal is to determine whether the volume of usable lumber can be estimated from the height of a tree. The results are plotted below.

Chapter 2



The variable _____ is the response variable in this study.
ANSWER: volume of usable lumber

8. A researcher measured the height (in feet) and the volume (in cubic feet) of usable lumber of 32 cherry trees. The goal is to determine whether the volume of usable lumber can be estimated from the height of a tree. The results are plotted below.



Select all descriptions that apply to the scatterplot.

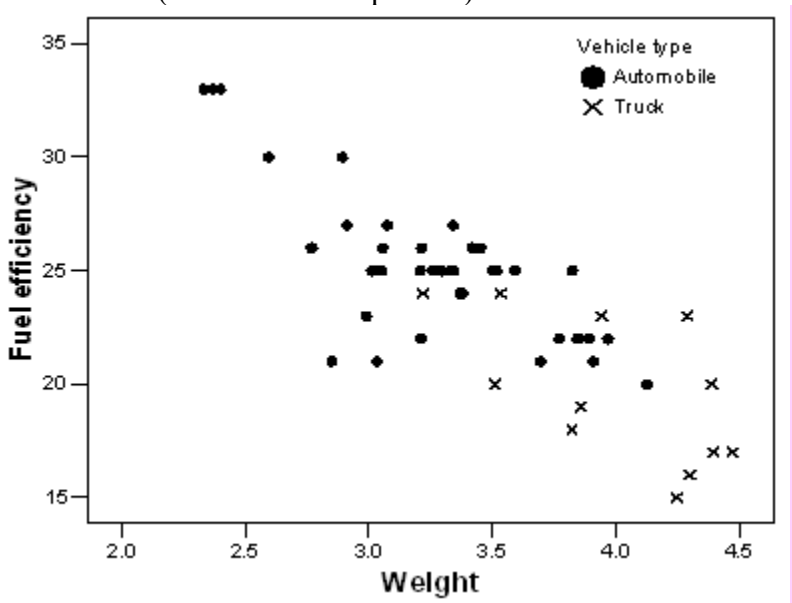
a. There is a positive association between height and volume.
b. There is a negative association between height and volume.
c. There is an outlier in the plot.
d. Both A and C.

ANSWER: d

Name: _____ Class: _____ Date: _____

Chapter 2

9. The graph below is a plot of the fuel efficiency (in miles per gallon, or mpg) of various cars versus the weight of these cars (in thousands of pounds).



The points denoted by the plotting symbol × correspond to pickup trucks and SUVs. The points denoted by the plotting symbol • correspond to automobiles (sedans and station wagons). From this plot, we can conclude:

a. only that there is little difference between trucks and automobiles.
b. only that trucks tend to be higher in weight than automobiles.
c. only that trucks tend to get poorer gas mileage than automobiles.
d. nothing, because the plot is invalid. A scatterplot is used to represent quantitative variables, and vehicle type is a qualitative variable.
e. Both B and C are correct.

ANSWER: e

10. In order for us to examine the relationship between two variables, the variables must be measured from the same _____.

a. cases
b. labels
c. units
d. values

ANSWER: a

11. Variables measured on the same cases are said to be _____ if knowing the values of one of the variables gives you information that you did not know before about the values of another variable.

a. transformed
b. categorical
c. associated

Chapter 2

d. quantitative

ANSWER: c

12. A variable that explains or causes change to another variable is called a(n) _____ variable.

a. independent
b. dependent
c. response

ANSWER: a

13. Two variables are _____ if knowing the values of one of the variables gives one information about the other variable.

a. associated
b. lurking
c. confounded

ANSWER: a

14. We are interested in determining whether students who graduate from larger universities receive greater starting salaries than students who graduate from smaller universities. We collected data from 50 small universities and 50 large universities to examine this relationship. This is an example of _____.

a. exploratory data analysis
b. benchmarking
c. data mining

ANSWER: b

15. A categorical variable can be added to a scatterplot.

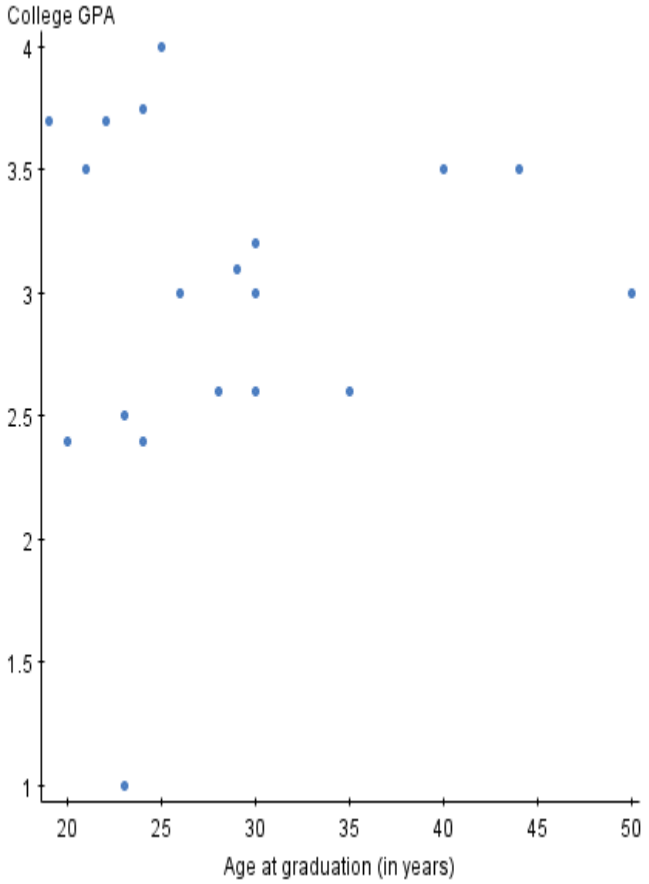
a. True
b. False

ANSWER: a

16. The scatterplot below displays data collected from 20 adults on their age and overall GPA at graduation.

Name: _____ Class: _____ Date: _____

Chapter 2



The scatterplot shows a strong relationship.

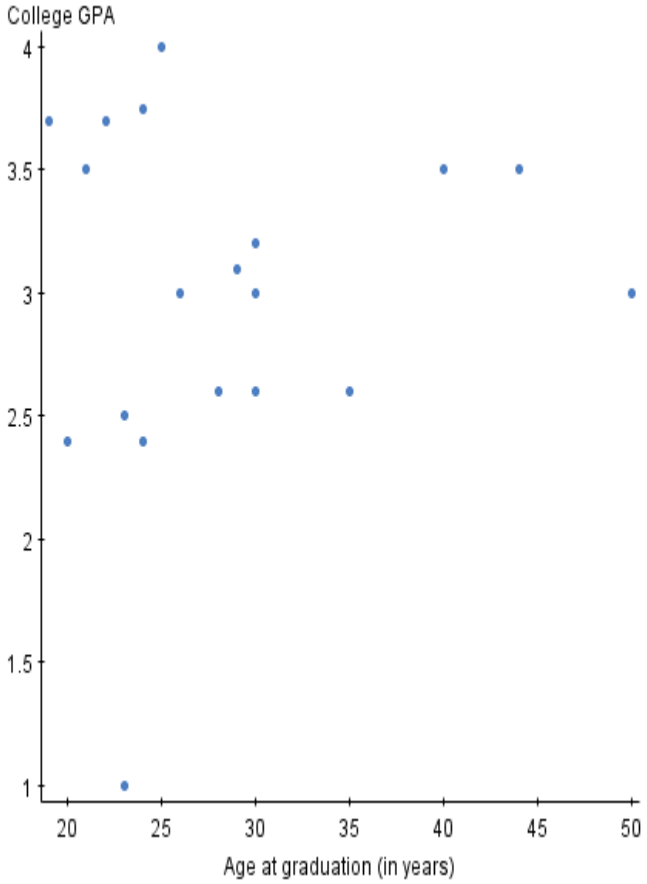
a. True

b. False

ANSWER: b

17. The scatterplot below displays data collected from 20 adults on their age and overall GPA at graduation.

Chapter 2



There appear to be outliers in the data set.

a. True
b. False

ANSWER: a

18. Which type of transformation may help change a curved relationship into a more linear relationship?

a. Log
b. Arcsine
c. Reciprocal
d. Cube root

ANSWER: a

19. Transformations are used to _____.

a. make curved relationships more linear
b. make a distribution more Normal
c. change the scale of measurements
d. All of the above

ANSWER: d

Chapter 2

20. In order for us to use a log transformation, all values must be positive.
    a. True
    b. False

ANSWER: a

21. The "direction" in scatterplots refers to the _____ directions.
    a. horizontal and vertical
    b. positive and negative
    c. left and right
    d. None of the above

ANSWER: b

22. Scatterplot "smoothing" is used to determine the _____ of the data.
    a. direction
    b. form
    c. variation
    d. None of the above

ANSWER: b

23. Scatterplots can be used to determine _____ relationships between variables.
    a. linear
    b. quadratic
    c. cubic
    d. All of the above
    e. None of the above

ANSWER: d

24. Which of the following statements about a scatterplot is(are) TRUE?
    a. It is always necessary to identify one of the two variables as the explanatory variable and the other as the response variable.
    b. On a scatterplot we look for overall patterns showing the form, direction, and shape of the relationship.
    c. Both A and B are true statements.
    d. None of the above statements is true.

ANSWER: b

25. Explanatory variables are also called _____ variables.

ANSWER: independent

26. Response variables are also called _____ variables.

ANSWER: dependent

Chapter 2

27. Time plots are special scatterplots where the explanatory variable, *x*, is a measure of time.
    a. True
    b. False

ANSWER: a

28. You can describe the overall pattern of a scatterplot by the _____.
    a. form, direction, and strength
    b. Normal distribution
    c. number of points in the plot
    d. None of the above

ANSWER: a

29. When examining a scatterplot for form, you are looking to see whether _____.
    a. the points in the scatterplot show a straight-line pattern
    b. the points in the scatterplot show a curved relationship
    c. there are clusters in the scatterplot
    d. None of the above
    e. A, B, and C

ANSWER: e

30. When examining a scatterplot for direction, you are looking to see if _____.
    a. high values of the two variables in the scatterplot tend to occur together
    b. high values of one variable tend to occur with low values of the other variable
    c. there is a positive association
    d. there is a negative association
    e. All of the above
    f. A and C only.
    g. B and D only.

ANSWER: e

31. When examining a scatterplot for strength, you are looking to see _____.
    a. how closely the points in the scatterplot follow a line
    b. how closely the points in the scatterplot follow a curve
    c. All of the above
    d. None of the above

ANSWER: c

32. When looking for relationships between two quantitative variables, you are looking for _____.
    a. linear relationships

Chapter 2

b. nonlinear relationships
c. Both of these
d. Neither of these

ANSWER: c

33. An outlier is _____.
a. a point in a scatterplot that follows the same pattern as the other points
b. a point in a scatterplot that does not follow the same pattern as the other points
c. Neither of these

ANSWER: b

34. Two variables are positively associated when _____.
a. above-average values of one tend to accompany above-average values of the other, and vice versa
b. above-average values of one tend to accompany below-average values of the other, and vice versa
c. both variables have an outlier
d. None of the above

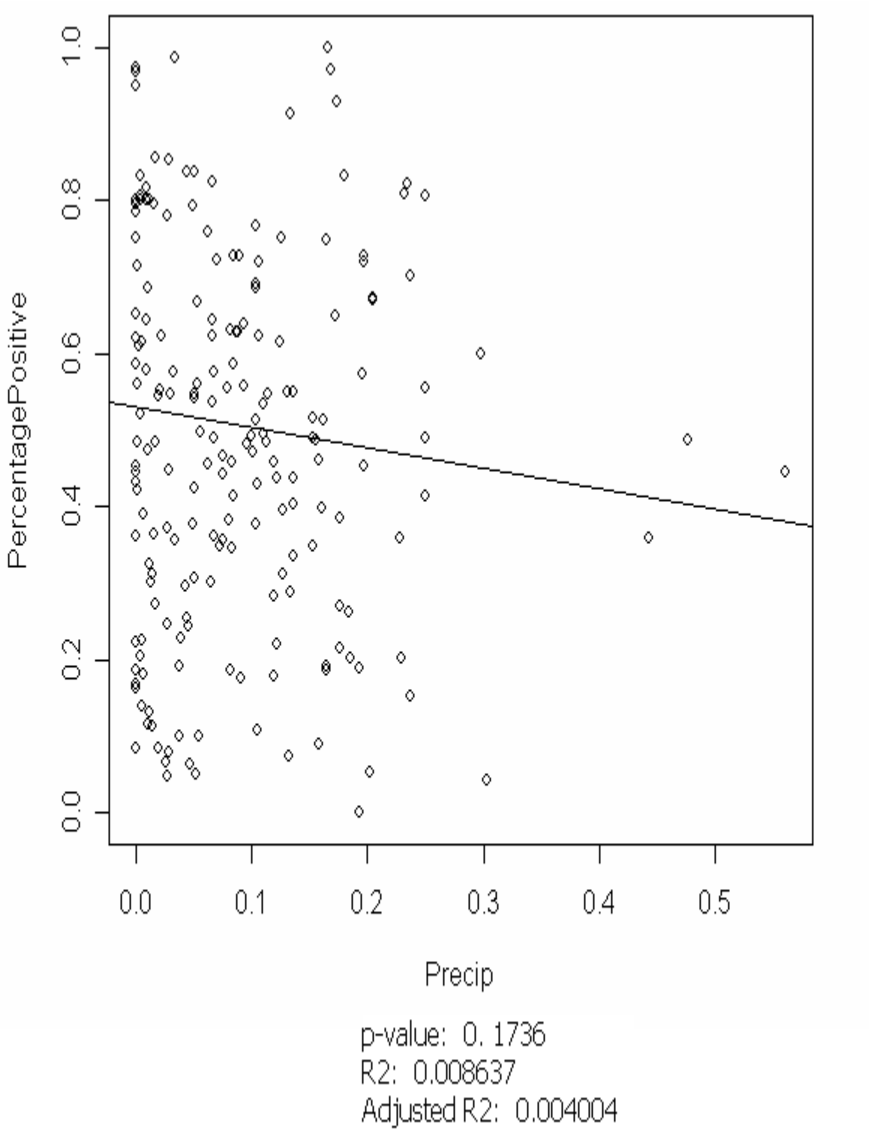
ANSWER: a

35. If you have two quantitative variables, one way to study them is to use a _____.
a. scatterplot
b. two-way table
c. Neither of the above

ANSWER: a

36. Malaria is a leading cause of infectious disease and death worldwide. It is also a familiar example of a vector-borne disease that could be greatly affected by the influence of climate change. The scatterplot shows total precipitation (in mm) in select cities in West Africa on the *x* axis, and the percent of people who tested positive for malaria in the select cities on the *y* axis, in the year 2000.

Chapter 2



There is a strong linear relationship between percent of people who tested positive for malaria and precipitation.

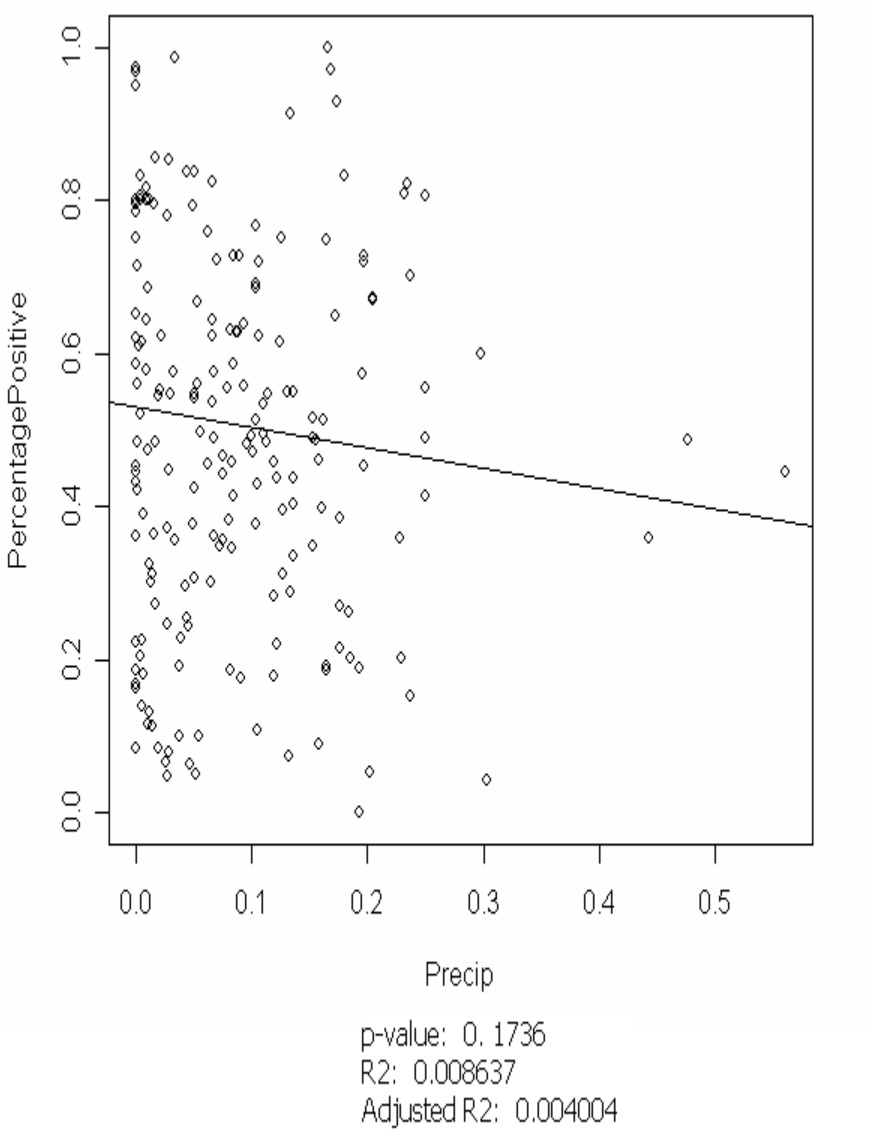
a. True

b. False

ANSWER: b

37. Malaria is a leading cause of infectious disease and death worldwide. It is also a familiar example of a vector-borne disease that could be greatly affected by the influence of climate change. The scatterplot shows the total precipitation (in mm) in select cities in West Africa on the *x* axis, and the percent of people who tested positive for malaria in the select cities on the *y* axis, in the year 2000.

Chapter 2



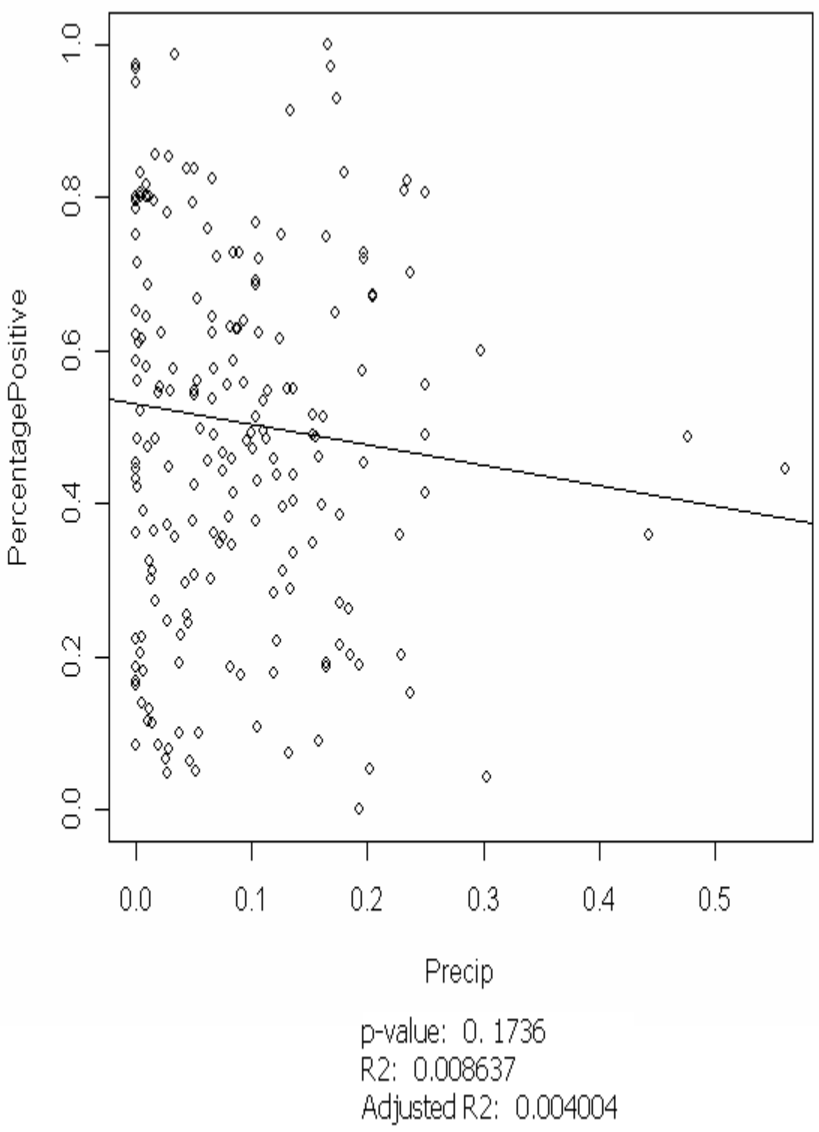
There are influential points in the scatterplot.

a. True
b. False

ANSWER: a

38. Malaria is a leading cause of infectious disease and death worldwide. It is also a familiar example of a vector-borne disease that could be greatly affected by the influence of climate change. The scatterplot shows the total precipitation (in mm) in select cities in West Africa on the *x* axis, and the percent of people who tested positive for malaria in the select cities on the *y* axis, in the year 2000.

Chapter 2



Precipitation is the _____ variable.

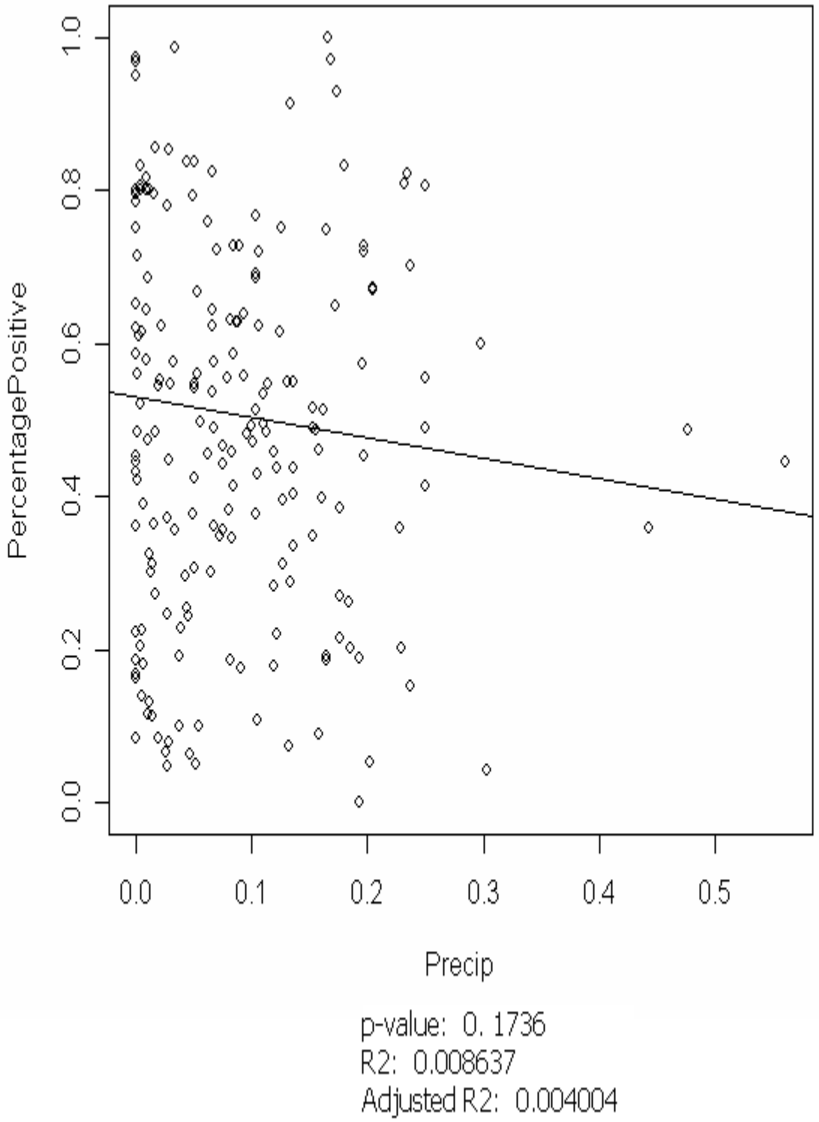
a. independent
b. dependent
c. response
d. explanatory
e. A and C
f. A and D

ANSWER: f

39. Malaria is a leading cause of infectious disease and death worldwide. It is also a familiar example of a vector-borne disease that could be greatly affected by the influence of climate change. The scatterplot shows the

Chapter 2

total precipitation (in mm) in select cities in West Africa on the *x* axis, and the percent of people who tested positive for malaria in the select cities on the *y* axis, in the year 2000.



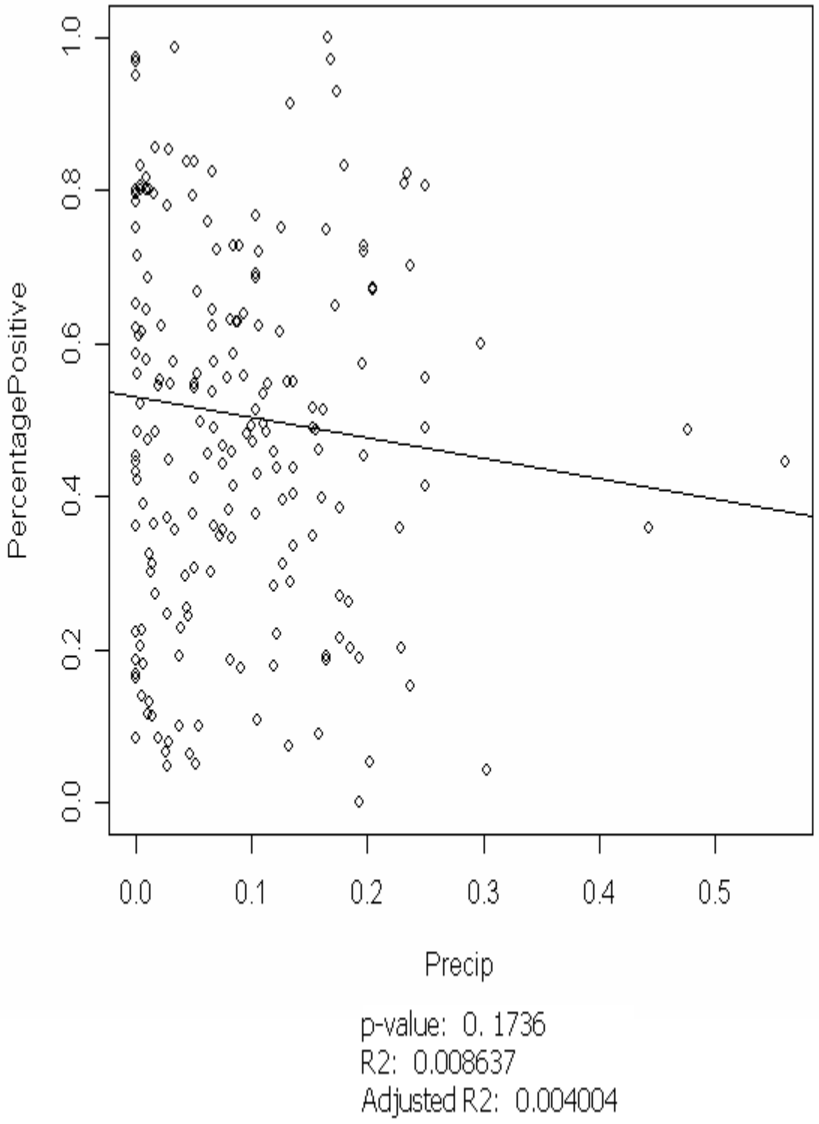
Percent who tested positive for malaria is the _____ variable.

a. independent
b. dependent
c. response
d. explanatory
e. B and C
f. B and E

ANSWER: e

Chapter 2

40. Malaria is a leading cause of infectious disease and death worldwide. It is also a familiar example of a vector-borne disease that could be greatly affected by the influence of climate change. The scatterplot shows the total precipitation (in mm) in select cities in West Africa on the *x* axis, and the percent of people who tested positive for malaria in the select cities on the *y* axis, in the year 2000.



The correlation between precipitation and percent who tested positive for malaria is probably close to _____.

a. 1
b. 0
c. impossible to determine

ANSWER: b

41. Which one of the following statements is true?

a. The correlation, *r*, measures the strength of the linear relationship between two quantitative

variables.

b. The correlation, *r*, measures the strength of the linear relationship between two categorical variables.

c. The correlation, *r*, measures the strength between one quantitative variable and one categorical variable.

ANSWER: a

42. The lack of a linear relationship between two quantitative variables is represented by the correlation, *r*, with values _____.

a. much greater than zero

b. much less than zero

c. approximately equal to zero

d. close to 1 or to –1 than to zero

ANSWER: c

43. A college newspaper interviews a psychologist about a proposed system for rating the teaching ability of faculty members. The psychologist says, "The evidence indicates that the correlation between a faculty member's research productivity and teaching rating is close to zero." What would be a correct interpretation of this statement?

a. Good researchers tend to be poor teachers, and vice versa.

b. Good teachers tend to be poor researchers, and vice versa.

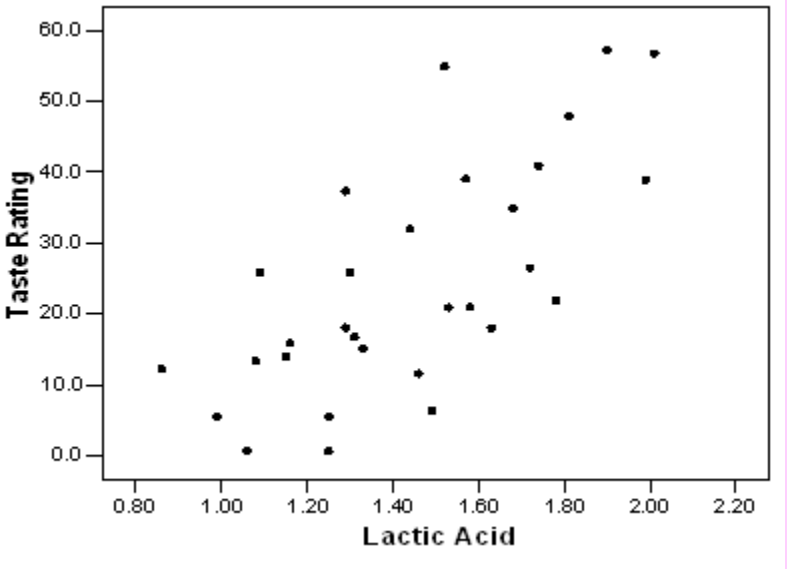
c. Good researchers are just as likely to be good teachers as they are to be bad teachers. Likewise for poor researchers.

d. Good research and good teaching go together.

ANSWER: c

44. As Swiss cheese matures, a variety of chemical processes take place. The taste of matured cheese is related to the concentration of several chemicals in the final product. In a study of cheese in a certain region of Switzerland, samples of cheese were analyzed for lactic acid concentration and were subjected to taste tests. The numerical taste scores were obtained by combining the scores from several tasters. A scatterplot of the observed data is shown below.

Chapter 2

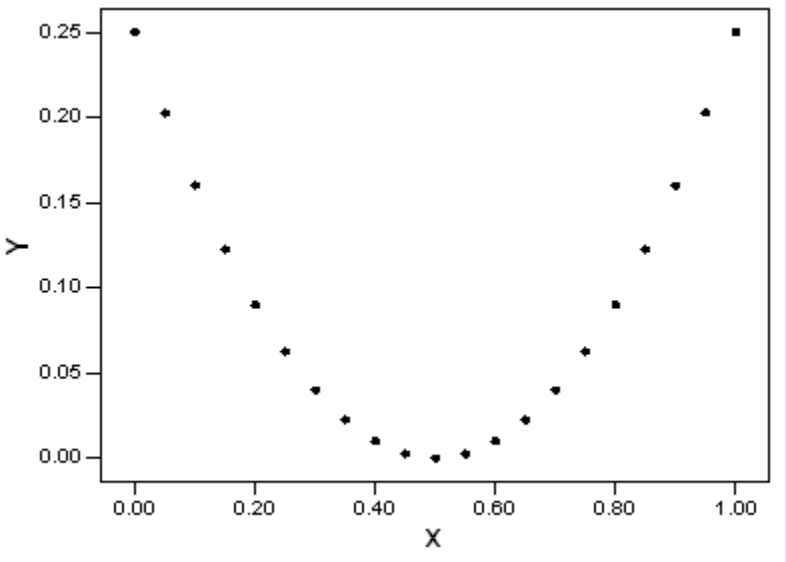


What is a plausible value for the correlation between lactic acid concentration and taste rating?

a. 0.999
b. 0.7
c. 0.07
d. –0.7

ANSWER: b

45. Consider the following scatterplot of two variables *x* and *y*.



What can we conclude from this graph?

a. The correlation between *x* and *y* must be close to 1 because there is a nearly perfect relationship between them.
b. The correlation between *x* and *y* must be close to –1 because there is a nearly perfect relationship between them, but it is not a straight-line relation.

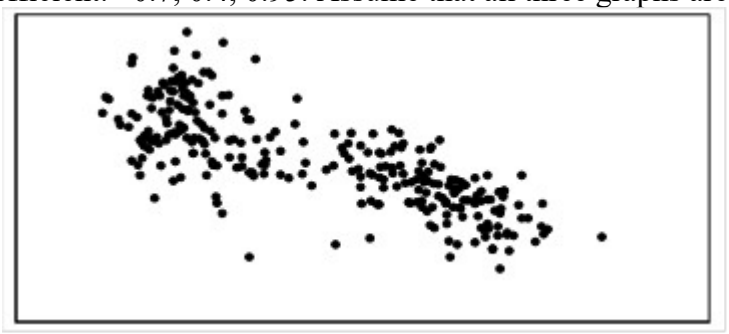
Chapter 2

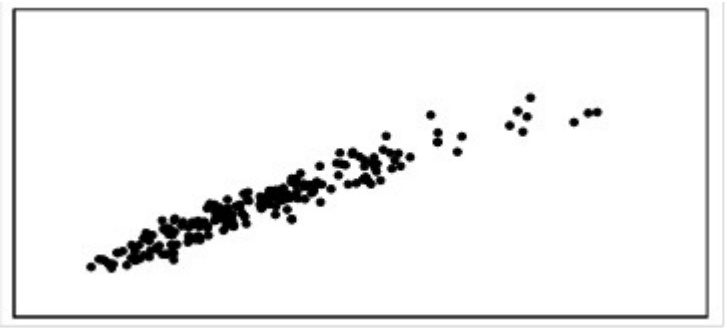
c. The correlation between x and y is close to 0.

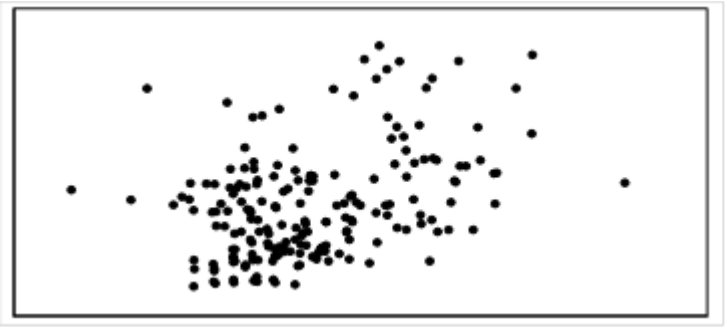
d. The correlation between x and y could be any number between –1 and +1.Without knowing the actual values, we can say nothing more.

ANSWER: c

46. Match the three graphs labeled A, B, and C with the following three possible values of the correlation coefficient: –0.7, 0.4, 0.95. Assume that all three graphs are made on the same scale.

A. 

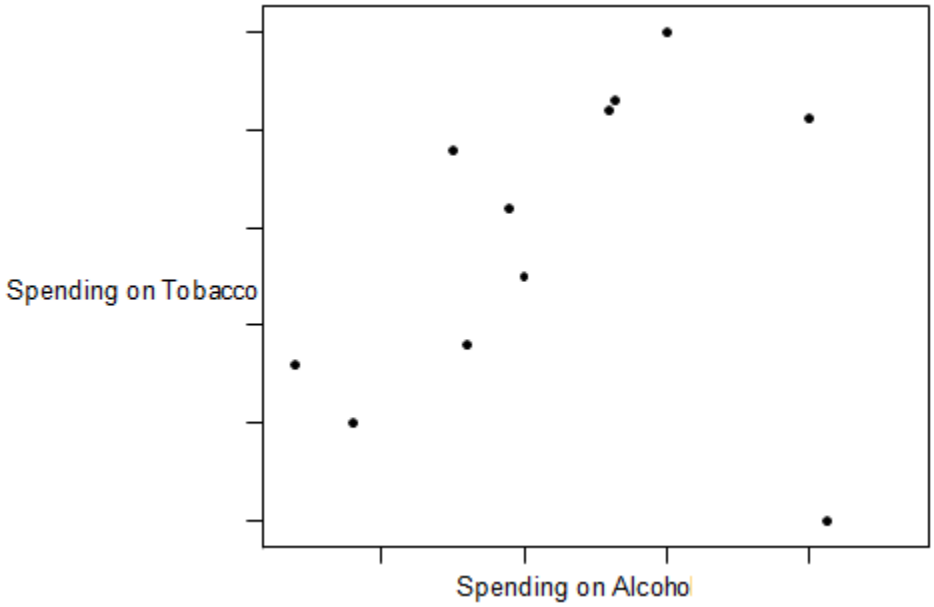
B. 

C. 

ANSWER: A. –0.7, B. 0.95, C. 0.4

47. The British government conducts regular surveys of household spending. The average weekly household spending on tobacco products and the average weekly household spending on alcoholic beverages for each of 11 regions in Great Britain were recorded. A scatterplot of spending on tobacco versus spending on alcohol is given below.

Chapter 2



Determine whether each of the following statements is true or false.
A. The observation in the lower-right corner of the plot is influential.
B. There is clear evidence of a negative association between spending on alcohol and spending on tobacco.
C. The equation of the least-squares regression line for this plot would be approximately $y = 10 - 2x$.
D. If we measured the spending in dollars instead of pounds, the correlation coefficient would decrease, because a dollar is worth less than a pound.
ANSWER: A. True, B. False, C. False, D. False

48. An experiment is conducted to study the bonding strength of adhesives that contain varying amounts of a particular chemical additive. Wafers of a specified material are glued together using the adhesive with each amount of additive and allowed to set for 24 hours, and then the strength needed to separate the wafers is determined. It is reported that the correlation between strength required and amount of additive was 0.86 pound-force per square inch.
Fill in the blanks: This report is _____ because correlation must be _____.

a. incorrect; unitless
b. correct; positive
c. incorrect; negative
d. None of the above

ANSWER: a

49. Which of the following statements best describes correlation?

a. Correlation measures whether the relationship between two quantitative variables is linear.
b. Correlation measures how much of the change in the response variable is caused by a change in the explanatory variable.
c. Correlation measures the strength of the relationship between any two variables.
d. Correlation measures the strength of the linear relationship between two quantitative variables.
e. Correlation measures the strength of the linear association between two categorical variables.

Chapter 2

ANSWER: d

50. Correlation is a measure of the direction and strength of the linear (straight-line) association between two quantitative variables. The analysis of data from a study found that the scatterplot between two variables, *x* and *y*, appeared to show a straight-line relationship, and the correlation was calculated to be –0.84. This tells us that:

a. there is little reason to believe that the two variables have a linear association relationship.
b. all of the data values for the two variables lie on a straight line.
c. there is a strong linear relationship between the two variables, with larger values of *x* tending to be associated with larger values of the *y* variable.
d. there is a strong linear relationship between *x* and *y*, with smaller *x*-values tending to be associated with larger values of the *y* variable.
e. there is a weak linear relationship between *x* and *y*, with smaller *x*-values tending to be associated with smaller values of the *y* variable.

ANSWER: d

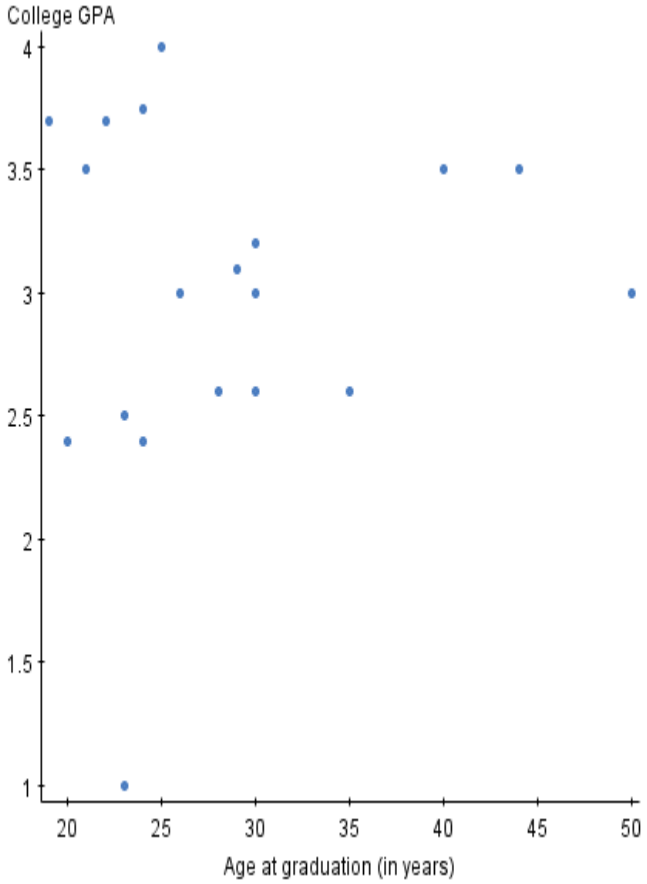
51. In a study of 1991 model cars, a researcher computed the least-squares regression line of price (in dollars) on horsepower. He obtained the following equation for this line.
Price = –6677 + 175 × horsepower
Based on the least-squares regression line, what would we predict to be the cost of a 1991 model car with horsepower equal to 200?

a. $41,677
b. $35,000
c. $28,323
d. $13,354

ANSWER: c

52. The scatterplot below displays data collected from 20 adults on their age and overall GPA at graduation.

Chapter 2



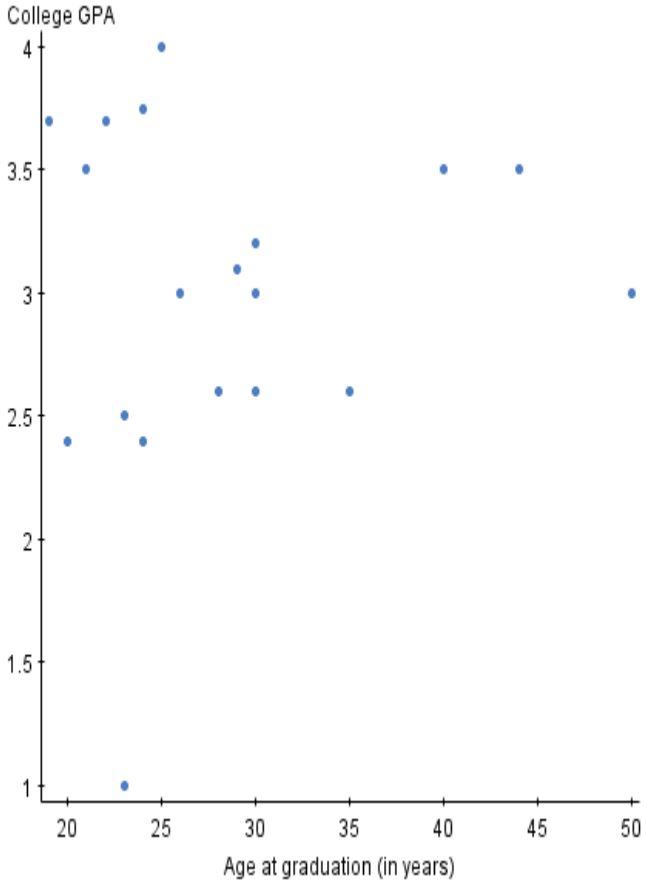
How would removing the outliers located at the points (23, 1.00) and (50, 3.00) affect the correlation, *r*?

a. *r* would not change.
b. *r* would probably increase.
c. *r* would probably decrease.

ANSWER: b

53. The scatterplot below displays data collected from 20 adults on their age and overall GPA at graduation.

Chapter 2



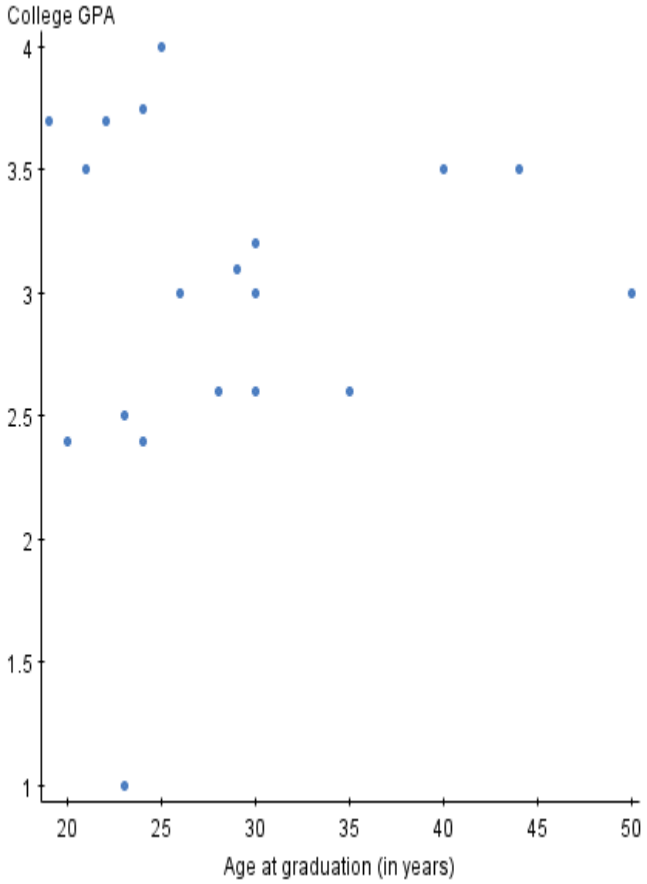
After removal of the outliers located at the points (23, 1.00) and (50, 3.00), the correlation, *r,* is likely to be ____.

a. positive
b. negative
c. zero
d. 1

ANSWER: a

54. The scatterplot below displays data collected from 20 adults on their age and overall GPA at graduation.

Chapter 2



The correlation, *r*, is likely to reveal _____ between the two variables.

a. a strong correlation
b. a weak correlation
c. no relationship
d. a curved relationship

ANSWER: b

55. Before using the correlation, *r*, you should do which of the following?

a. Look at the scatterplot of the data to determine whether the relationship appears linear.
b. Look at a histogram to be sure your data are approximately normal.
c. Look at a stem plot to determine whether the data are symmetric.
d. All of the above

ANSWER: a

56. The measurement units for the correlation, *r*, are determined from _____.

a. the variable on the *x* axis
b. the variable on the *y* axis
c. either the variable on the *x* axis or the variable on the *y* axis
d. None of the above.

Chapter 2

ANSWER: d

57. Which plot helps you visualize the value of the correlation, *r*?

a. Histogram
b. Box plot
c. Scatterplot
d. Density curve

ANSWER: c

58. Suppose you are examining the correlation between two quantitative variables, and the correlation, *r*, is very small. However, you expected it to be larger. What could you do?

a. Examine the data to determine whether there are any outliers that could be removed. If so, remove the outliers and recalculate *r*.
b. Change the units of measurement to something else (for example, convert data measured in inches to centimeters).
c. Plot the data on a smaller scale.
d. None of the above

ANSWER: a

59. The value of r^2 ranges between:

a. –1 and 1.
b. 0 and 1.
c. 0 and 100.
d. –100 and 100.

ANSWER: b

60. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis from the software *StatCrunch*.

Simple linear regression results:
Dependent Variable: Mortality.rate
Independent Variable: Obesity.rate
Mortality.rate = 13.458199 – 0.21749489 Obesity.rate
Sample size: 3098
R (correlation coefficient) = –0.0067
R-sq = 4.5304943E-5
Estimate of error standard deviation: 111.20661

Name: _____ Class: _____ Date: _____

Chapter 2

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	–0.21749489	0.5807189	≠ 0	3096	–0.37452698	0.708

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	*P*-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3097	3.8289688E7			

What fraction of the variation in mortality rates is explained by the least-squares regression?

a. 0
b. 1
c. –0.0067
d. 13.45

ANSWER: a

61. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis from the software *StatCrunch*.

Simple linear regression results:
Dependent Variable: Mortality.rate
Independent Variable: Obesity.rate
Mortality.rate = 13.458199 – 0.21749489 Obesity.rate
Sample size: 3098
R (correlation coefficient) = –0.0067
R-sq = 4.5304943E-5
Estimate of error standard deviation: 111.20661

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	–0.21749489	0.5807189	≠ 0	3096	–0.37452698	0.708

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	*P*-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3097	3.8289688E7			

Name: _____ Class: _____ Date: _____

Chapter 2

What is the equation to predict mortality rates from obesity rates?

a. Mortality.rate = 13.458199 – 0.21749489 Obesity.rate
b. Obesity.rate = 13.458199 – 0.21749489 Mortality.rate
c. Mortality.rate = 13.458199 + 0.21749489 Obesity.rate
d. Mortality.rate = 13.458199 – 0.0067 Obesity.rate

ANSWER: a

62. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis from the software *StatCrunch*.

Simple linear regression results:
Dependent Variable: Mortality.rate
Independent Variable: Obesity.rate
Mortality.rate = 13.458199 – 0.21749489 Obesity.rate
Sample size: 3098
R (correlation coefficient) = –0.0067
R-sq = 4.5304943E-5
Estimate of error standard deviation: 111.20661

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	−0.217494 9	0.5807189	≠ 0	3096	–0.37452698	0.708

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	*P*-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3097	3.8289688E7			

The correlation between obesity rate and CRC mortality rate is _____.

a. very strong
b. very weak
c. moderately strong
d. moderately weak

ANSWER: b

63. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis

Chapter 2

from the software *StatCrunch*.
Simple linear regression results:
Dependent Variable: Mortality.rate
Independent Variable: Obesity.rate
Mortality.rate = 13.458199 – 0.21749489 Obesity.rate
Sample size: 3098
R (correlation coefficient) = –0.0067
R-sq = 4.5304943E-5
Estimate of error standard deviation: 111.20661

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	–0.21749489	0.5807189	≠ 0	3096	–0.37452698	0.708

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3097	3.8289688E7			

The explanatory variable is _____.

a. obesity rate
b. CRC mortality rate
c. slope
d. intercept

ANSWER: a

64. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis from the software *StatCrunch*.
Simple linear regression results:
Dependent Variable: Mortality.rate
Independent Variable: Obesity.rate
Mortality.rate = 13.458199 – 0.21749489 Obesity.rate
Sample size: 3098
R (correlation coefficient) = –0.0067
R-sq = 4.5304943E-5
Estimate of error standard deviation: 111.20661

Chapter 2

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	–0.21749489	0.5807189	≠ 0	3096	–0.37452698	0.708

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3 9 7	3.8289688E7			

The response variable is _____.

a. obesity rate
b. CRC mortality rate
c. slope
d. intercept

ANSWER: b

65. The least-squares regression line always passes through the point ____.

a. (0, 0)
b. $(\bar{x},\bar{y})$
c. (median of *x*, median of *y*)
d. None of the above

ANSWER: b

66. The explanatory variable and the response variable can be interchanged in regression, just as in correlations.

a. True
b. False

ANSWER: b

67. Least-squares regression can be used for prediction between explanatory and response variables that have a _____ relationship.

a. linear
b. quadratic
c. cubic
d. All of the above

ANSWER: a

68. Before performing a least-squares regression analysis, one should _____.

Chapter 2

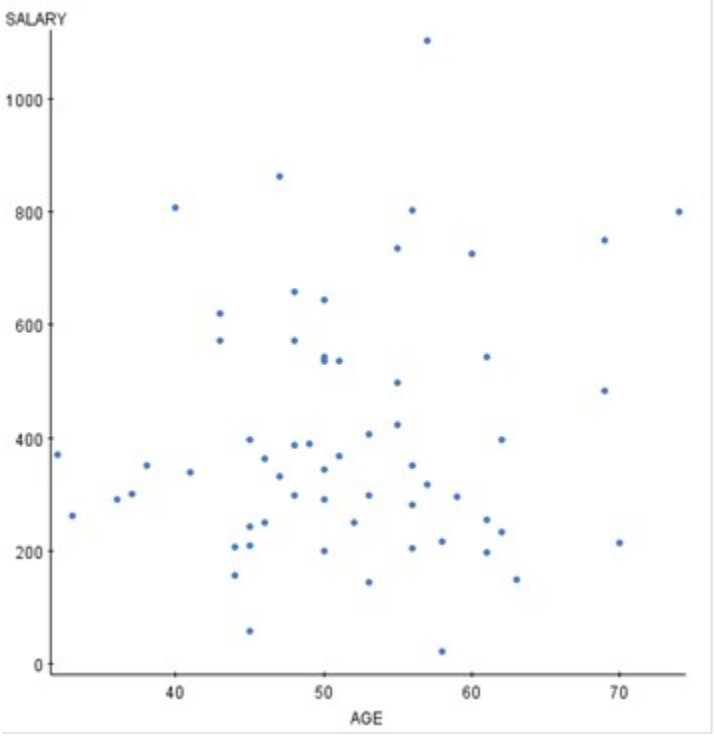
a. examine a scatterplot of the data to look for the type of relationship between the data.
b. examine the data for possible outliers
c. make sure the explanatory variable has a Normal distribution
d. All of the above
e. Only A and B

ANSWER: e

69. The least-squares regression line is the line that _____.
a. makes the sum of the squares of the vertical distance of the data points from the line as small as possible.
b. makes the sum of the squares of the horizontal distance of the data points from the line as small as possible.
c. makes the sum of the squares of the vertical distance of the data points from the line as large as possible.
d. makes the sum of the squares of the vertical distance of the data points from the line zero.

ANSWER: a

70. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 59

Chapter 2

R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246

Analysis of variance table for regression model:

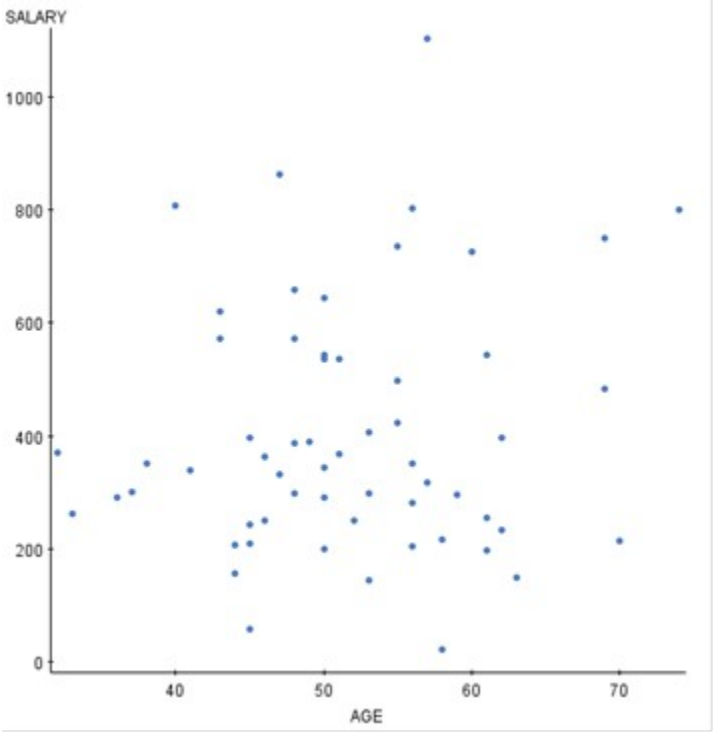
Source	DF	SS	MS	F-stat	*P*-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

Is age a good predictor of salary?

a. Yes, the intercept is high.
b. Yes, the correlation is high.
c. No, the intercept is too low.
d. No, the correlation and r^2 are low.

ANSWER: d

71. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 59
R (correlation coefficient) = 0.1276

Chapter 2

R-sq = 0.016270384
Estimate of error standard deviation: 220.64246

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	242.70212	168.7604	≠ 0	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	≠ 0	57	0.9709536	0.3357

Analysis of variance table for regression model:

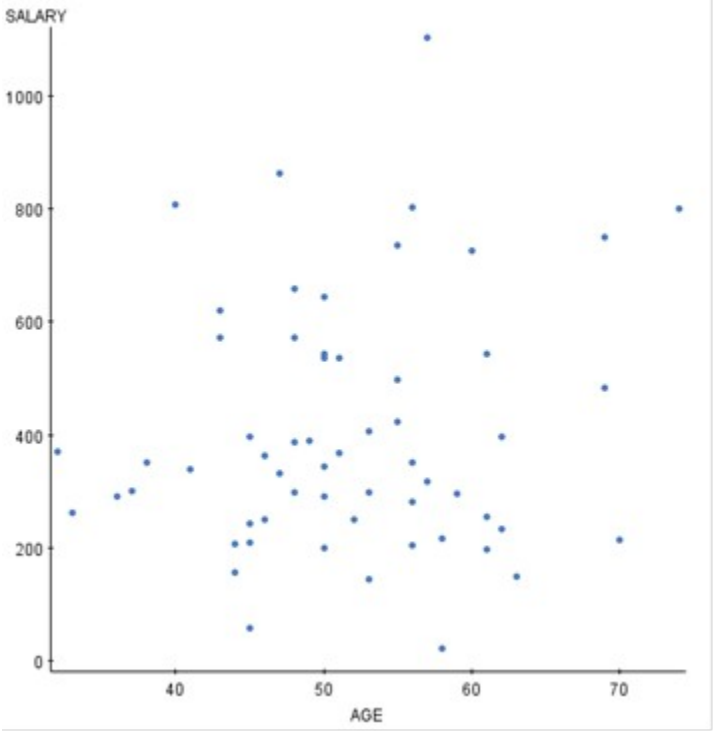
Source	DF	SS	MS	F-stat	*P*-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

Suppose a CEO is 57 years old. What do you predict his or her salary to be?

a. Over $400,000
b. Between $100,00 and $400,000
c. Under $100,000
d. None of the above

ANSWER: a

72. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.



Name: _____ Class: _____ Date: _____

Chapter 2

Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 59
R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	242.70212	168.7604	≠ 0	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	≠ 0	57	0.9709536	0.3357

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	*P*-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

Suppose you wanted to predict the salary of Facebook CEO Mark Zuckerberg, based on the information here. How well do you think your prediction would be, assuming Mr. Zuckerberg was 23 when he started Facebook and became CEO?

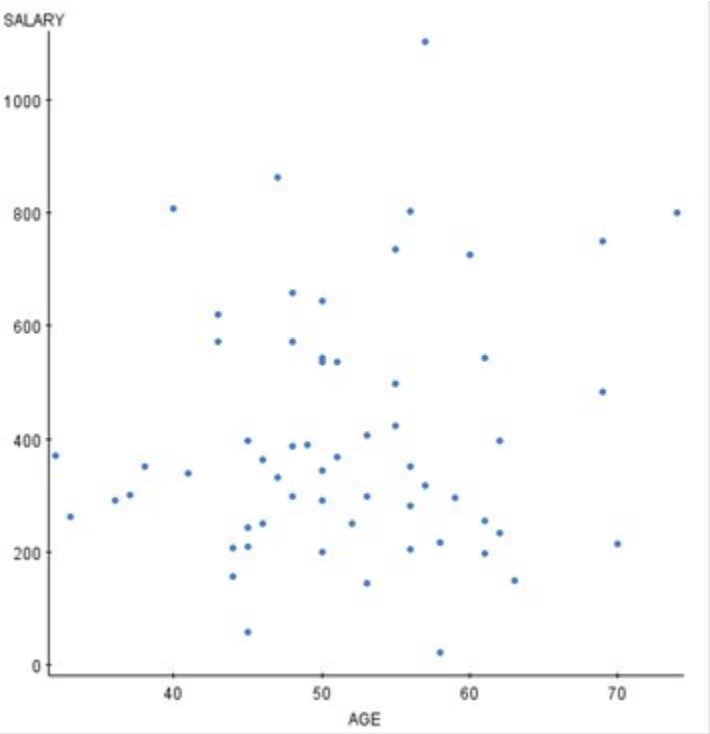
a. The prediction would be accurate and around $300,000.
b. The prediction would require extrapolation and therefore would not be accurate.
c. The prediction would be accurate and around $240,000.
d. None of the above

ANSWER: b

73. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.

Name: _____ Class: _____ Date: _____

Chapter 2



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 59
R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	242.70212	168.7604	≠ 0	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	≠ 0	57	0.9709536	0.3357

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	*P*-value
M del	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

How do you interpret the intercept for this problem?

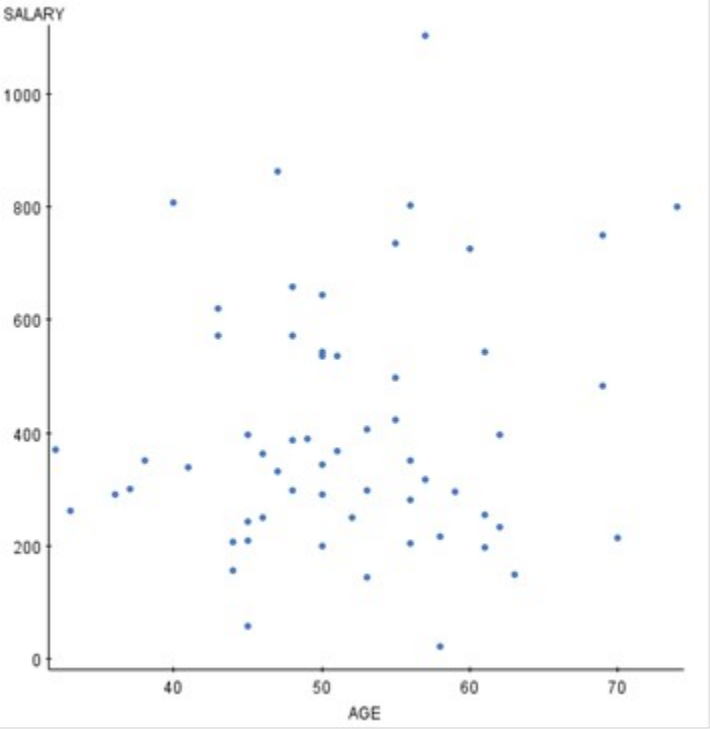
a. A CEO at birth is expected to make around $242,000.
b. A CEO at 100 years old is expected to make around $555.

Chapter 2

c. The intercept is not useful for this problem.
d. None of the above

ANSWER: c

74. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 59
R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	242.70212	168.7604	≠ 0	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	≠ 0	57	0.9709536	0.3357

Analysis of variance table for regression model:

Chapter 2

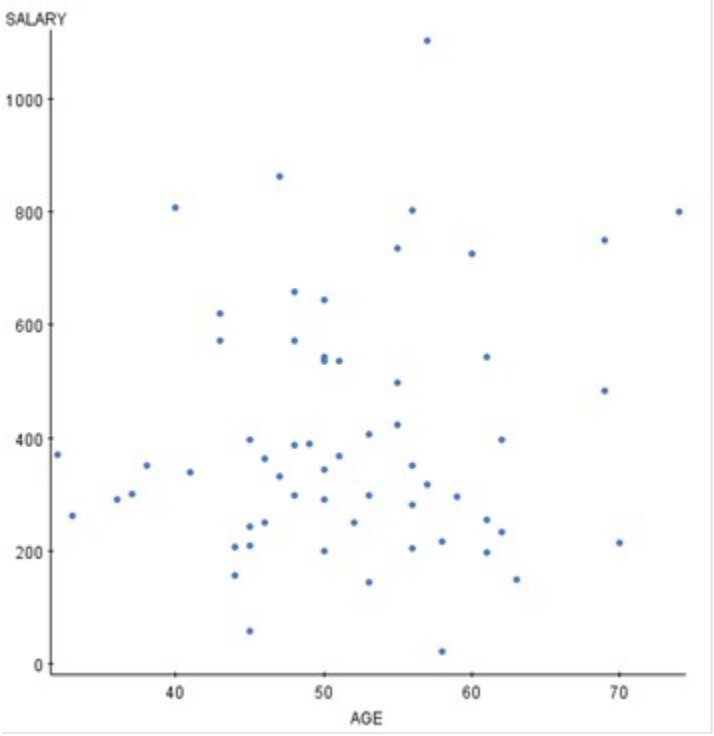
Source	DF	SS	MS	F-stat	*P*-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

After observing the scatterplot, one would expect the correlation to be:

a. strong.
b. weak.
c. close to 1.
d. None of the above

ANSWER: b

75. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 59
R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246
Parameter estimates:

Chapter 2

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	242.70212	168.7604	≠ 0	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	≠ 0	57	0.9709536	0.3357

Analysis of variance table for regression model:

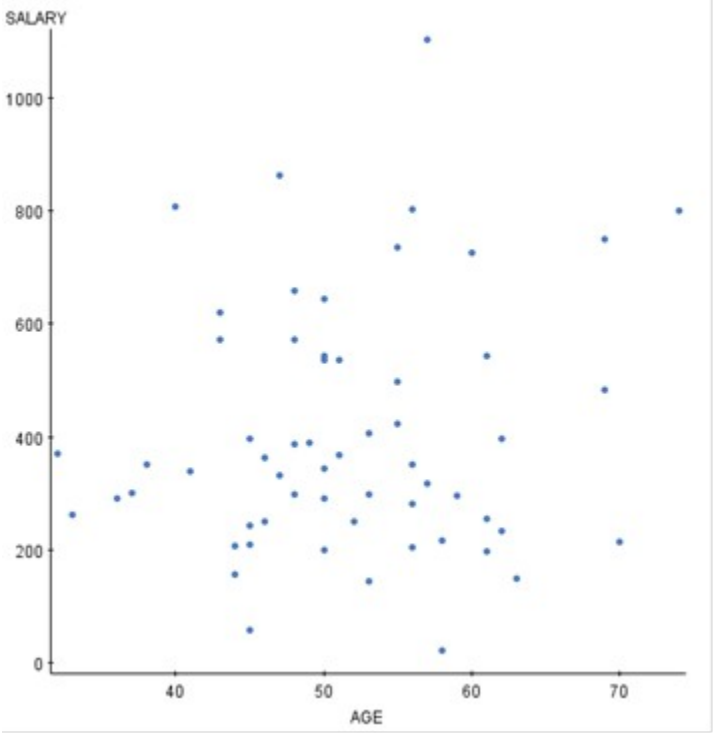
Source	DF	SS	MS	F-stat	*P*-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

What is one plausible reason for a correlation around 0.13 for this problem?

a. Age is a very strong predictor of CEO salary.
b. Age is not a good predictor, and something else may be a better predictor.
c. There are not enough data to accurately estimate the correlation.
d. The range of ages is too small.

ANSWER: b

76. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE

Chapter 2

Sample size: 59
R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246
Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	*P*-Value
Intercept	242.70212	168.7604	≠ 0	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	≠ 0	57	0.9709536	0.3357

Analysis of variance table for regression model:

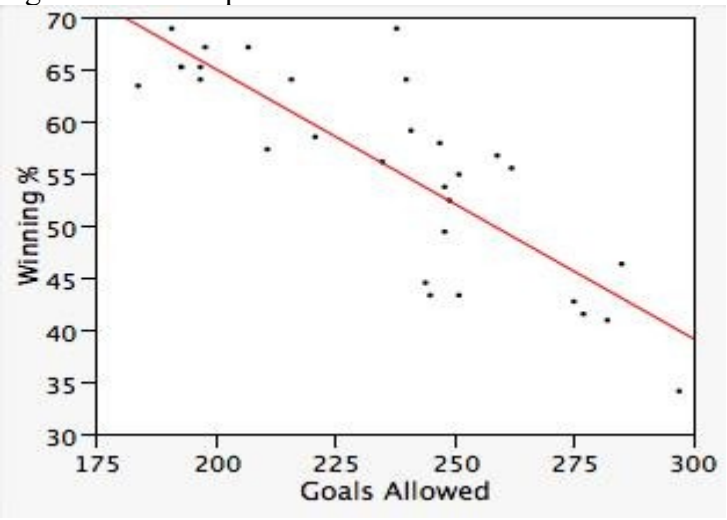
Source	DF	SS	MS	F-stat	*P*-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

What percent of the variation in CEO salaries is explained by age alone?

a. Around 1.6%
b. Around 0.016%
c. Around 0.12%
d. Around 12%

ANSWER: a

77. In the National Hockey League, a good predictor of the percent of games won by a team is the number of goals the team allows during the season. Data were gathered for all 30 teams in the NHL, and the scatterplot of their Winning Percent against the number of Goals Allowed in the 2006/2007 season with a fitted least-squares regression line is provided.



The least-squares regression line was calculated to be

Winning Percent (%) = 116.95 – 0.26 * Goals Allowed with $r^2 = 0.69$

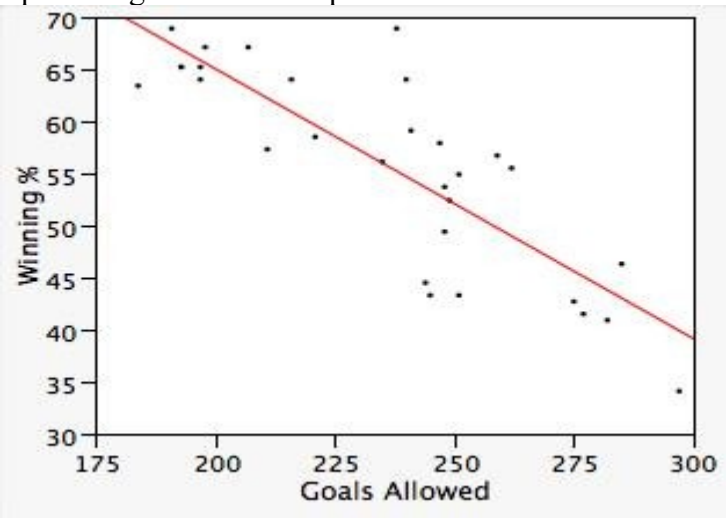
Which of the following provides the best interpretation of the slope of the regression line?

Chapter 2

a. If the Winning Percent increases by 1%, then the number of Goals Allowed decreases by 0.26.
b. If a team were to allow 100 goals during the season, its Winning Percent would be 90.95%.
c. If Goals Allowed increases by one goal, the Winning Percent increases by 0.26%.
d. If the Winning Percent increases by 1%, then the number of Goals Allowed increases by 0.26%.
e. If Goals Allowed increases by one goal, the Winning Percent decreases by 0.26%.

ANSWER: e

78. In the National Hockey League, a good predictor of the percent of games won by a team is the number of goals the team allows during the season. Data were gathered for all 30 teams in the NHL, and the scatterplot of their **Winning Percent** against the number of **Goals Allowed** in the 2006/2007 season with a fitted least-squares regression line is provided.



The least-squares regression line was calculated to be

Winning Percent (%) = 116.95 – 0.26 * Goals Allowed with $r^2 = 0.69$

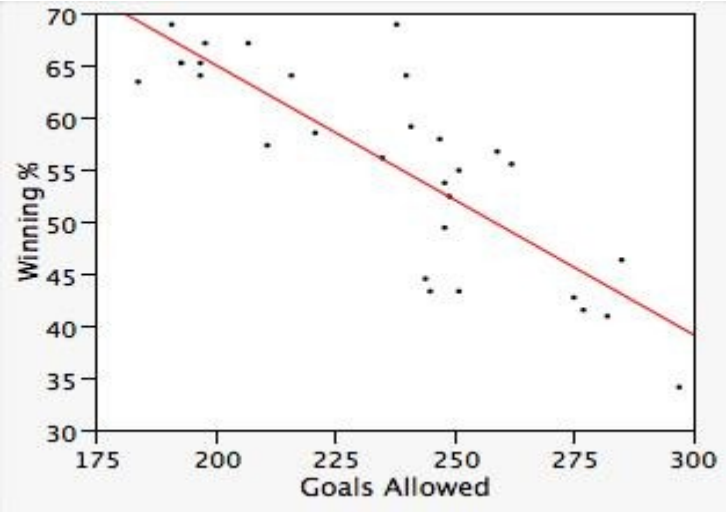
The Montréal Canadiens team allowed 251 goals in 2006/2007. Using the least-squares regression line, the prediction of the team's Winning Percent would be _____%.

ANSWER: 51.69

79. In the National Hockey League, a good predictor of the percent of games won by a team is the number of goals the team allows during the season. Data were gathered for all 30 teams in the NHL, and the scatterplot of their **Winning Percent** against the number of **Goals Allowed** in the 2006/2007 season with a fitted least-squares regression line is provided.

Name: _____ Class: _____ Date: _____

Chapter 2



The least-squares regression line was calculated to be

Winning Percent (%) = 116.95 – 0.26 * Goals Allowed with $r^2 = 0.69$

For the Winning Percent and Goals Allowed least-squares regression analysis above, which of the following statements is(are) TRUE?

a. About 69% of the variation in the variable Goals Allowed can be explained by the least-squares regression of Winning Percent on Goals Allowed.

b. About 69% of the variation in the variable Winning Percent can be explained by the least-squares regression of Winning Percent on Goals Allowed.

c. If the correlation between Winning Percent and Goals Allowed were calculated, it would be 0.83.

d. A and C are true.

e. B and C are true.

ANSWER: e

80. In a statistics course, a linear regression equation was computed to predict the final exam score from the score on the midterm exam. The equation of the least-squares regression line was $y = 10 + 0.9x$, where y represents the final exam score and x is the midterm exam score. Suppose Joe scores a 90 on the midterm exam. What would be the predicted value of his score on the final exam?

a. 81

b. 89

c. 91

d. Cannot be determined from the information given. We also need to know the correlation.

ANSWER: c

81. 81. John's parents recorded his height at various ages between 36 and 66 months. Below is a record of the results:

Age (months)	36	48	54	60	66
Height (inches)	34	38	41	43	45

John's parents decide to use the least-squares regression line of John's height on age to predict his height at age

Chapter 2

21 years (252 months). What conclusion can we draw?

a. John's height, in inches, should be about half his age, in months.
b. The parents will get a fairly accurate estimate of his height at age 21 years, because the data are clearly correlated.
c. Such a prediction could be misleading because it involves extrapolation.
d. All of the above

ANSWER: c

82. The correlation coefficient between two variables x and y is $r = 0.121$. What conclusion can we draw?

a. Because the correlation is so low, the relationship between x and y is not very strong, so there is no use in studying this relationship.
b. Because the correlation is so low, we know that the linear relationship between x and y is not very strong, but there may be a different relationship between the two variables. We need to first look at a scatterplot.
c. The correlation between x and y is low, but that does not matter. We can still use least-squares regression to calculate an equation of the form $y = ax + b$.
d. None of the above

ANSWER: b

83. Many high school students take either the SAT or the ACT. However, some students take both. Data were collected from 60 students who took both college entrance exams. The average SAT score was 912 with a standard deviation of 180. The average ACT score was 21 with a standard deviation of 5. The correlation between the two variables equals 0.817.
To predict a student's SAT score from that student's ACT score, what is the equation of the least-squares regression line?

a. $= 0.3027 + 0.0227x$
b. $= 294.348 + 29.412x$
c. $= 156 + 36x$
d. Cannot be determined from the information given

ANSWER: b

84. Many high school students take either the SAT or the ACT. However, some students take both. Data were collected from 60 students who took both college entrance exams. The average SAT score was 912 with a standard deviation of 180. The average ACT score was 21 with a standard deviation of 5. The correlation between the two variables equals 0.817.
What fraction of the variation in the values of the SAT scores is accounted for by the linear relationship between SAT and ACT scores?

a. 66.7%
b. 81.7%
c. 90.4%
d. Cannot be determined from the information given

ANSWER: a

Chapter 2

85. Recall that when we standardize the values of a variable, the standardized value has a mean of 0 and a standard deviation of 1. Suppose we measure two variables x and y on each of several subjects. We standardize both variables and compute the least-squares regression line of y on x for these standardized values. Suppose the slope of this least-squares regression line is –0.44. What conclusion can we draw?

a. The intercept will be 1.0.
b. The intercept will also be –0.44.
c. The correlation will be 1.0.
d. The correlation will be –0.44.

ANSWER: d

86. A researcher at a large company has collected data on the beginning salary and current salary of 48 randomly selected employees. The least-squares regression equation for predicting their current salary from their beginning salary is $y = –7 + 2.12x$. The current salaries had a mean of $32,070 with a standard deviation of $15,300. The beginning salaries had a mean of $16,340 with a standard deviation of $5,970. What is the correlation between current salary and beginning salary?

a. $r = 0.390$
b. $r = 0.506$
c. $r = 0.827$
d. Cannot be determined from the information given

ANSWER: c

87. A researcher at a large company has collected data on the beginning salary and current salary of 48 randomly selected employees. The least-squares regression equation for predicting their current salary from their beginning salary is $y = –2532.7 + 2.12x$. Joseph Keller started working for the company earning $22,000. What do you predict his current salary to be?

a. $39,560.22
b. $44,107.30
c. $46,640.00
d. $49,172.70

ANSWER: b

88. A researcher at a large company has collected data on the beginning salary and current salary of 48 randomly selected employees. The least-squares regression equation for predicting their current salary from their beginning salary is $y = –2532.7 + 2.12x$. Kathy Jones started working for the company earning $19,000. She currently earns $40,000. What is the *residual* for Ms. Jones?

a. $1187.30
b. $2252.70
c. $2812.70
d. Cannot be determined from the information given

ANSWER: b

Chapter 2

89. Which of the following statements about least-squares regression involving two quantitative variables, x and y, is(are) TRUE?

a. A change of 1 standard deviation in x corresponds to a change of r standard deviations in y.

b. The least-squares regression line always passes through the point $(\bar{x},\bar{y})$.

c. The square of the correlation, r^2, is the fraction of the variation in the values of y that is explained by the least-squares regression of y on x.

d. The least-squares regression line of y on x is the line that makes the sum of the squares of the vertical distances of the data points from the line as small as possible.

e. All of the above are true.

ANSWER: e

90. Data were obtained from the A&W website about the total fat (in grams) and the protein content (in grams) for various items on the menu. Some summary statistics are provided.

Item	**Total fat (grams)**	**Protein (grams)**
Kid's Cheeseburger	24	23
Kid's Hamburger	22	21
Original Bacon Cheeseburger	33	27
Original Bacon Double Cheeseburger	48	45
Original Double Cheeseburger	42	40
Papa Burger	42	41

	Total fat (grams)	**Protein (grams)**
Mean	35.167	32.833
Standard Deviation	10.591	10.362

Correlation $r = 0.983$

The slope of the least-squares regression line for total fat on protein is _____.

a. –0.998
b. 1.005
c. 0.962
d. 2.170
e. 0.966

ANSWER: b

91. Data were obtained from the A&W website about the total fat (in grams) and the protein content (in grams) for various items on its menu. Some summary statistics are provided.

Item	**Total fat (grams)**	**Protein (grams)**
Kid's Cheeseburger	24	23
Kid's Hamburger	22	21
Original Bacon Cheeseburger	33	27
Original Bacon Double Cheeseburger	48	45
Original Double Cheeseburger	42	40
Papa Burger	42	41

	Total fat (grams)	**Protein (grams)**

Chapter 2

Mean	35.167	32.833
Standard Deviation	10.591	10.362

Correlation $r = 0.983$

The intercept for the least-squares regression line of total fat on protein is _____.

a. –0.998
b. 1.005
c. 0.962
d. 2.170
e. 0.966

ANSWER: d

92. Data were obtained from the A&W website about the total fat (in grams) and the protein content (in grams) for various items on its menu. Some summary statistics are provided.

Item	**Total fat (grams)**	**Protein (grams)**
Kid's Cheeseburger	24	23
Kid's Hamburger	22	21
Original Bacon Cheeseburger	33	27
Original Bacon Double Cheeseburger	48	45
Original Double Cheeseburger	42	40
Papa Burger	42	41

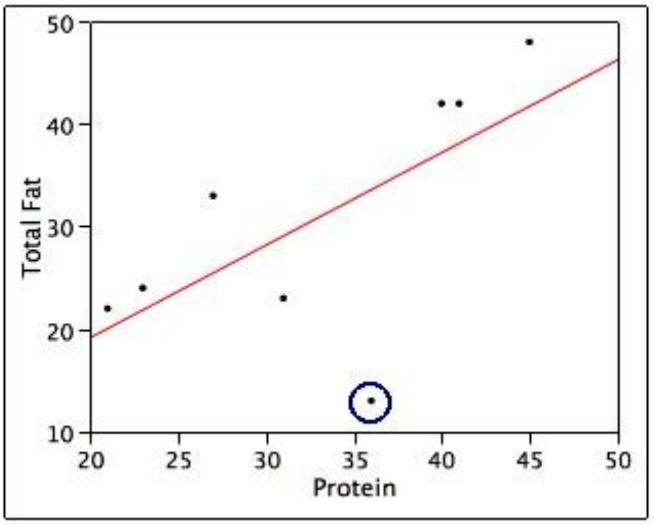
	Total fat (grams)	**Protein (grams)**
Mean	35.167	32.833
Standard Deviation	10.591	10.362

Correlation $r = 0.983$

Additional data on total fat and protein were found for two additional A&W menu items. These were as follows:

Item	**Total fat (grams)**	**Protein (grams)**
Crispy Chicken Sandwich	23	31
Grilled Chicken Sandwich	13	36

The scatterplot for the set of eight A&W menu items follows.



Chapter 2

The circled data point on the scatterplot is for the Grilled Chicken Sandwich.
Which of the following statements about the circled data point on the scatterplot is(are) TRUE?

a. This point would probably be considered an outlier.
b. The residual associated with this data point will have a negative value.
c. This point may be considered influential, but that depends on how much it affects the plot of the residuals.
d. Only A and B are true.
e. A, B, and C are true.

ANSWER: d

93. Using least-squares regression, it is determined that the logarithm (base 10) of the population of a country is related to the year by the following equation:
log(population) = –13.5 + 0.01 × (year)
Based on this equation, what will the (approximate) population of the country in the year 2016 be?

a. 6.56
b. 706
c. 2,006,000
d. 3,630,780
e. None of the above

ANSWER: e

94. A(n) ____ is an observation that is substantially different from the other observations.

a. outlier
b. lurking variable
c. confounding variable
d. None of the above

ANSWER: a

95. Correlation and regression are resistant to outliers.

a. True
b. False

ANSWER: b

96. 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
Amount of eroded soil = 0.4 + 1.3 × (flow rate)
What do we know about the correlation between amount of eroded soil and flow rate?

a. $r = 1/1.3$
b. $r = 0.4$

Chapter 2

c. It would be positive, but we cannot determine the exact value.

d. It could be either positive or negative. It is impossible to say anything about the correlation from the information given.

ANSWER: c

97. 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

Amount of eroded soil = 0.4 + 1.3 × (flow rate)

One of the flow rates used by the researcher was 0.3 liter per second, and for this flow rate the amount of eroded soil was 0.8 kilogram. These values were used in the calculation of the least-squares regression line. What is the residual corresponding to these values?

a. 0.01
b. –0.01
c. 0.5
d. –0.5

ANSWER: a

98. Researchers studied a sample of 100 adults between the ages of 25 and 35 and found a strong negative correlation between the amount of vitamin C an individual consumed and the number of pounds the individual was overweight. Which of the following can we conclude?

a. This is strong but not conclusive evidence that large amounts of vitamin C inhibit weight gain.

b. If the amount of vitamin C consumed and the number of pounds overweight for each individual in this study were plotted on a scatterplot, the points would lie close to a negatively sloping straight line.

c. If a larger sample of adults between the ages of 25 and 35 had been studied, the correlation would have been even stronger.

d. All of the above

ANSWER: b

99. The least-squares regression line is fit to a set of data. One of the data points has a positive residual. Determine whether each of the following statements is true or false.

A. The correlation between the values of the response and explanatory variables must be positive.
B. The point must lie above the least-squares regression line.
C. The point must lie near the right edge of the scatterplot.
D. The point must be influential.

ANSWER: A. False, B. True, C. False, D. False

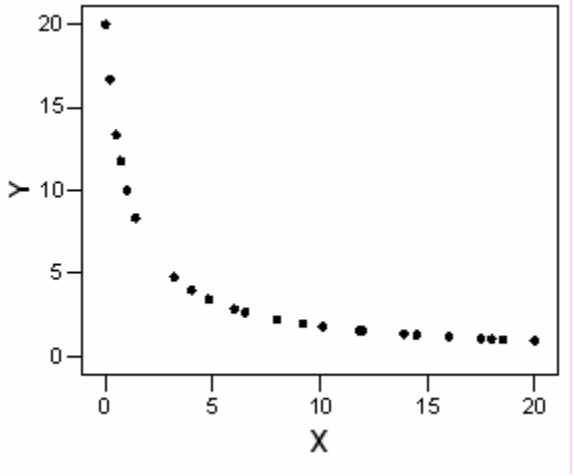
100. Determine whether each of the following statements regarding residuals is true or false.

A. The sum of the residuals is always 0.

Chapter 2

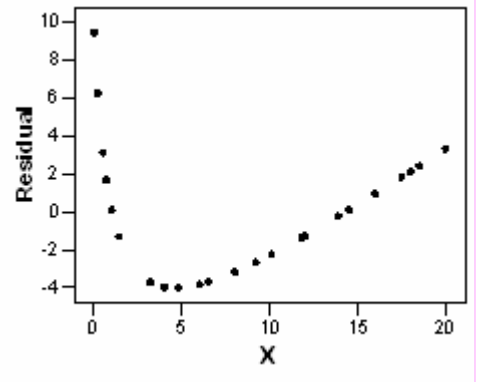
B. A plot of the residuals is useful for assessing the fit of the least-squares regression line.
C. The value of a residual is the observed value of the response minus the value of the response that one would predict from the least-squares regression line.
ANSWER: A. True, B. True, C. True

101. A response variable *y* and an explanatory variable *x* were measured on each of several subjects. A scatterplot of the measurements is given.

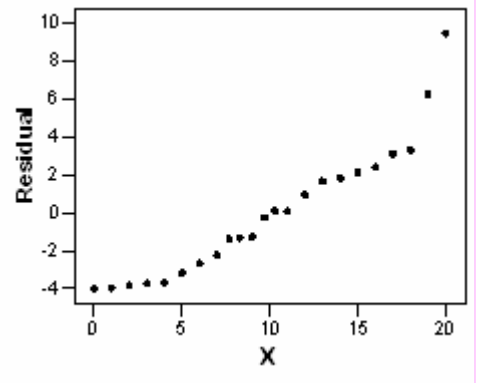


Which of the following is a plot of the residuals for the above data versus *x*?

a.

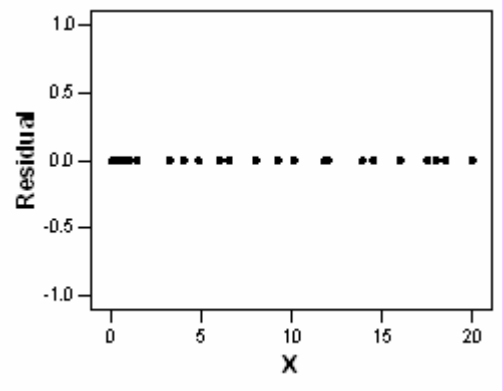


b.

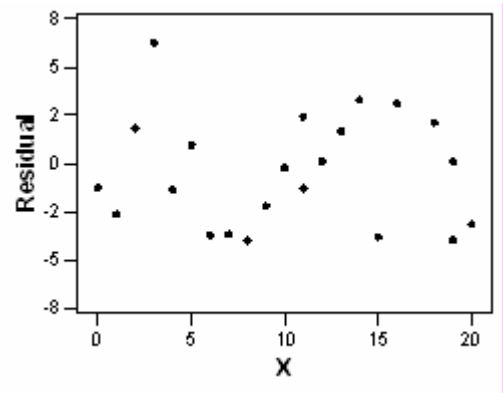


Chapter 2

c.



d.



ANSWER: a

102. A researcher studies the relationship between "Math SAT score plus Verbal SAT score" and "grade point average (GPA)" for college students at the end of their freshman year. In order to use a relatively homogeneous group of students, the researcher examines only data of high school valedictorians (students who graduated at the top of their high school class) who have completed their first year of college. The researcher finds the correlation between total SAT score and GPA at the end of the freshman year to be very close to 0. Which of the following would be a valid conclusion from these facts?

a. Because the group of students studied is homogeneous, the results should give an accurate estimate of the correlation the researcher would find if all college students who had completed their freshman year were studied.
b. If the researcher had studied all college students who had completed their freshman year, the correlation would have been even smaller than that actually found by the researcher. By restricting the study to valedictorians, the researcher is examining a group that will be more informative than those students who have completed only their freshman year.
c. The researcher made a mistake. Correlation cannot be calculated (the formula for correlation is invalid) unless all students who completed their freshman year are included.
d. None of the above

ANSWER: d

103. When one is exploring very large sets of data involving many variables, which of the following statements is true?

a. Extrapolation is safe because it is based on a greater quantity of evidence.

Chapter 2

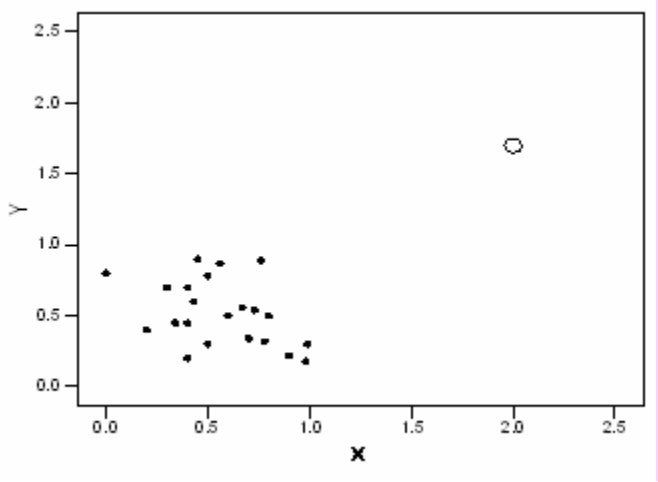
b. Associations will be stronger than would be seen in a much smaller subset of the data.

c. A strong association is good evidence for causation because it is based on a large quantity of information.

d. None of the above

ANSWER: d

104. Consider the scatterplot below.



What do we call the point indicated by the plotting symbol O?

a. A residual

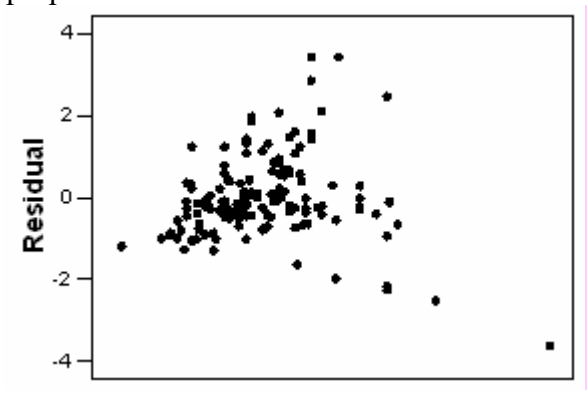
b. Influential

c. A *z*-score

ANSWER: b

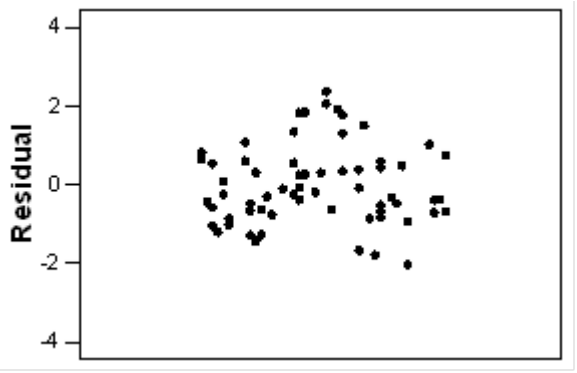
105. Four different residual plots are shown below. Which plots indicate that the linear model is not appropriate?

i)

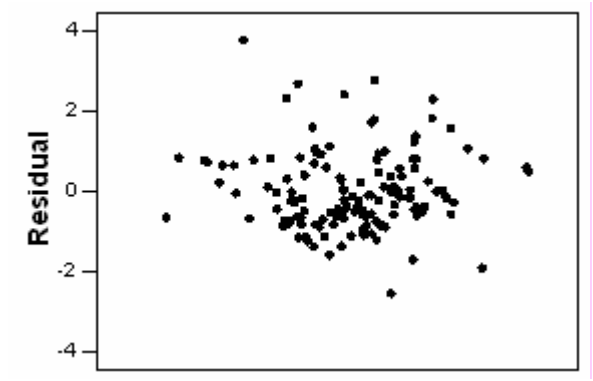


Chapter 2

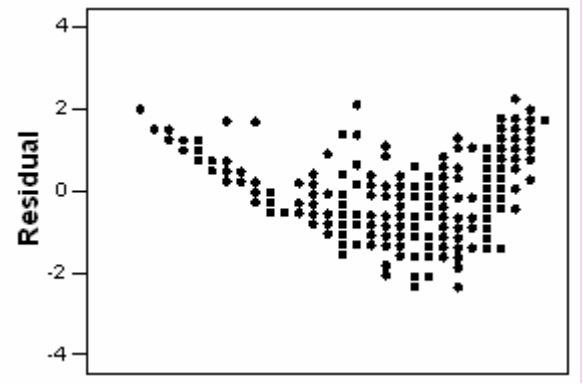
ii)



iii)



iv)



a. i and iii
b. iii and iv
c. i and iv
d. ii and iii

ANSWER: c

106. Plots of the residuals versus fits should show a linear pattern if the regression line is a good fit for your data.

a. True
b. False

ANSWER: b

107. Fill in the blank. Influential outliers are usually in the_____.

Chapter 2

a. *x* direction on a scatterplot
b. *y* direction on a scatterplot

ANSWER: a

108. Influential outliers are easy to detect, because the residuals will always be very large compared with the residuals of the other observations.

a. True
b. False

ANSWER: b

109. An electronics store is handing out a survey to its clients who buy a smartphone. Some of the questions on the survey ask the clients to rate the smartphone on ease of use, appearance, price, etc. Another question asks for the client's age. From all the different ratings on the survey, a total assessment score is calculated. The correlation between this total assessment score and the age of the client is –0.165. The store owner can legitimately conclude which of the following?

a. Older clients seem to not like smartphones.
b. There is a negative linear relationship between age and assessment score.
c. Age does not help much in predicting assessment score.
d. None of the above. We really need to look at a scatterplot of the data first.

ANSWER: d

110. Exploring extremely large data sets in hopes of finding patterns is called _____.

a. exploratory data analysis
b. extrapolation
c. data mining
d. None of the above

ANSWER: c

111. Correlations based on averages tend to be _____ correlations based on individuals.

a. higher than
b. lower than
c. the same as

ANSWER: a

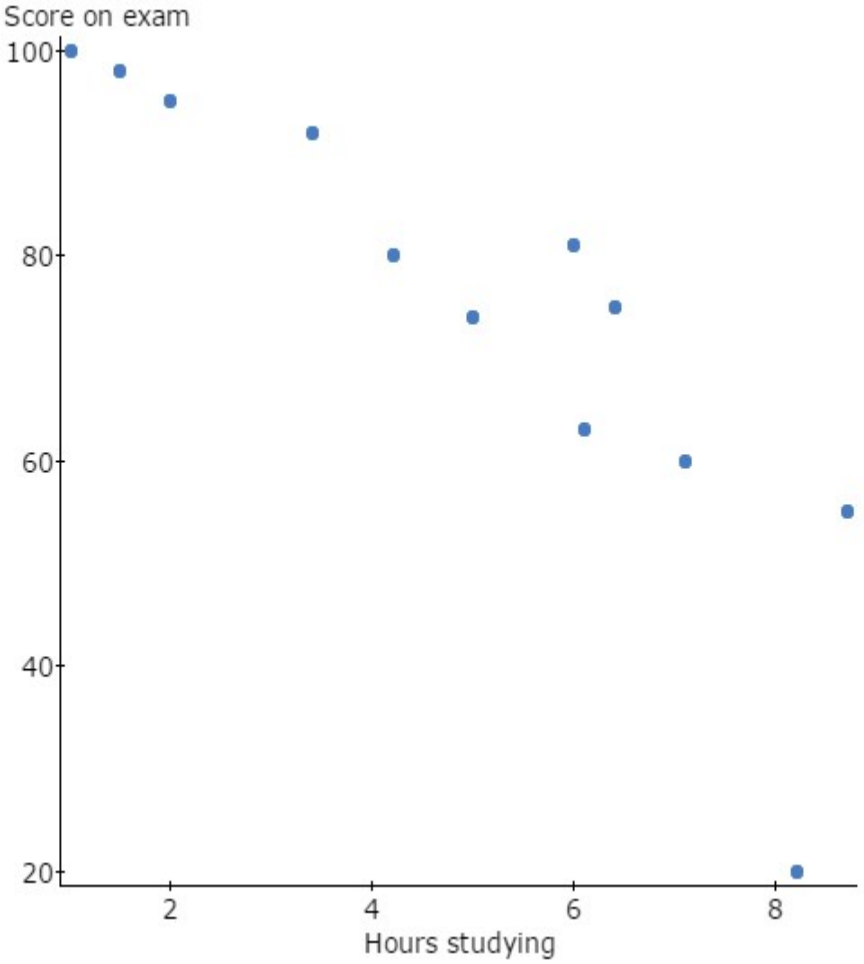
112. It is known that not exercising may lead to poor health. However, it is possible that people who are already in poor health do not have the ability or energy to exercise. This is an example of _____.

a. causation
b. common response
c. confounding
d. None of the above

ANSWER: c

Chapter 2

113. The scatterplot illustrates data from a basic statistics class. Students in the class were asked to provide the amount of time (in hours) they spent studying for the first exam. The professor then made a scatterplot to present the relationship between the number of hours a student studied and the score (from 0–100, with 100 being the best score) that the student received on the first exam. How would you interpret this scatterplot?



a. Students who studied the least amount of time received the highest grades. Therefore, they should not study long on a statistics exam if they want to receive a high grade.

b. Students who studied the most received the highest grades. Therefore, they should study several hours to receive the highest exam scores.

c. The correlation is likely a nonsense correlation caused by a lurking variable. Students who received higher scores probably did not need to study as much because they were doing better in the course than students who received lower scores.

d. None of the above

ANSWER: c

114. Correlations caused by lurking variables are sometimes called _____.

a. nonsense correlations

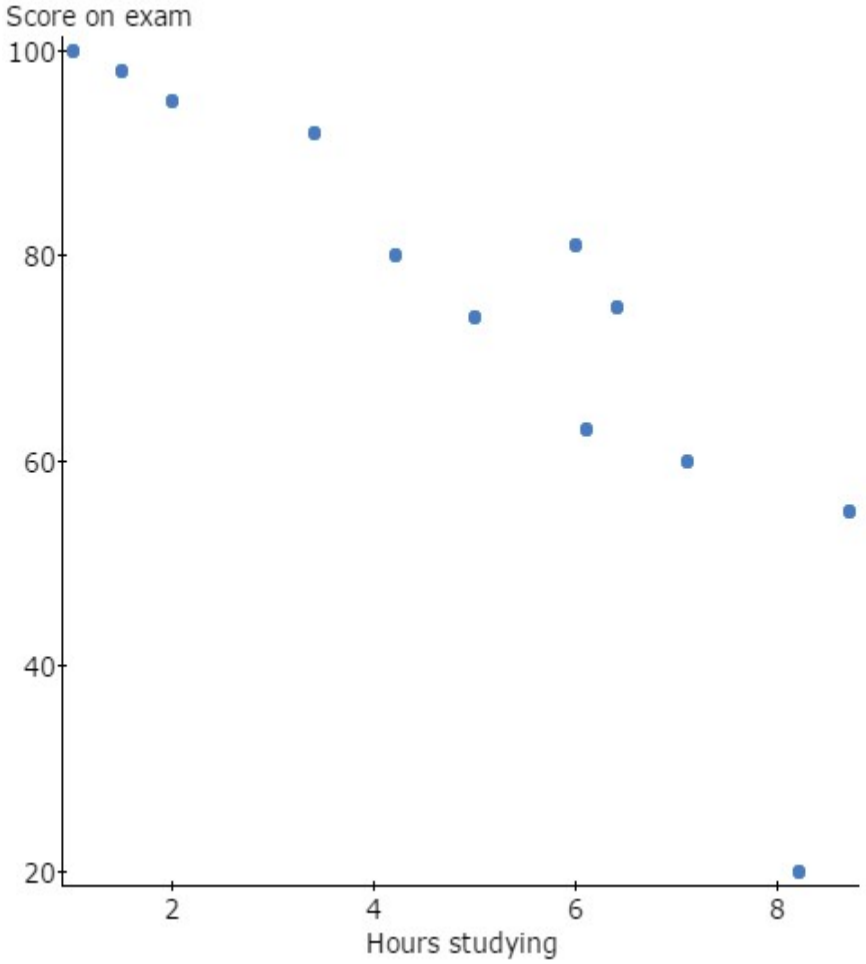
b. association correlations

Chapter 2

c. reverse correlations
d. None of the above

ANSWER: a

115. Give an example of a lurking variable that might explain the nonsense correlation between time spent studying for an exam and grade received on the exam based on the scatterplot.



a. The lurking variable might be "Current grade in the class." Students performing well in the class may not need to study long for exams.
b. The lurking variable might be "Study hours." The longer students study for the exam, the lower the grade they will receive on the exam.
c. The lurking variable might be "The exam." Students should be given different exams based on the amount of time they spent studying.
d. There are no lurking variables. Students should not study long for exams if they want to receive a high grade.

ANSWER: a

116. Finding patterns in truly large databases, such as tracking all Google searches over a year from everyone who used the search engine, requires the use of _____.

Chapter 2

a. exploratory data analysis (EDA)
b. regression analysis
c. data mining
d. None of the above

ANSWER: c

117. Data mining requires the use of_____.
a. efficient algorithms
b. the field of computer science
c. automated tools that can produce results from vague queries
d. All of the above

ANSWER: d

118. A study of the salaries of full professors at a small university shows that the median salary for female professors is considerably less than the median salary for male professors. Further investigation shows that the median salaries for male and female full professors are about the same in every department (English, physics, etc.) of the university. Which phenomenon explains the reversal in this example?
a. Extrapolation
b. Simpson's paradox
c. Causation
d. Correlation

ANSWER: b

119. 119. The California Department of State Police keeps track of the number of points received for various traffic violations by drivers. The department is interested in examining the relationship between the number of points received (low, medium, high) and the insurance premium. Some information on the point category and the insurance premium category is given.

Insurance premium category	Low	Medium	High
Cheap	12%	38%	50%
Medium	29%	33%	27%
Expensive	59%	29%	23%

Which distribution is displayed in the above table?
a. The joint distribution of premium category and point category
b. The marginal distribution of point category
c. The conditional distribution of premium category given point category
d. The conditional distribution of point category given premium category

ANSWER: c

120. The 94 students in a statistics class are categorized by gender and by year in school. The numbers obtained are displayed below.
Year in School

Chapter 2

Gender	Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1	2	9	17	2	31
Female	23	17	13	7	3	63
Total	24	19	22	24	5	94

What proportion of the statistics students in this class are sophomores?

a. 0.105
b. 0.202
c. 0.302
d. 19

ANSWER: b

121. The 94 students in a statistics class are categorized by gender and by year in school. The numbers obtained are displayed below.

	Year in School					
Gender	Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1	2	9	17	2	31
Female	23	17	13	7	3	63
Total	24	19	22	24	5	94

What proportion of the statistics students in this class are male?

a. 0.065
b. 0.105
c. 0.33
d. 31

ANSWER: c

122. The 94 students in a statistics class are categorized by gender and by year in school. The numbers obtained are displayed below.

	Year in School					
Gender	Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1	2	9	17	2	31
Female	23	17	13	7	3	63
Total	24	19	22	24	5	94

The data are going to be summarized by computing the conditional distributions of year in school for male and female students. What would be the entry for male sophomores?

a. 0.065
b. 0.105
c. 0.33
d. 2

ANSWER: a

123. A student organization is trying to decide whether or not to offer more movies on campus. The

organization wants to determine whether this idea will appeal to members of both genders. A random sample of 1000 students were asked whether they were in favor of more movies on campus. The results by gender are shown in the table below.

	Opinion		
Gender	In favor	No opinion	Opposed
Male	330	165	55
Female	225	180	45

What proportion of the sampled students are in favor of more movies on campus?

a. 0.33
b. 0.5
c. 0.555
d. 0.6

ANSWER: c

124. A student organization is trying to decide whether or not to offer more movies on campus. The organization wants to determine whether this idea will appeal to members of both genders. A random sample of 1000 students were asked whether they were in favor of more movies on campus. The results by gender are shown in the table below.

	Opinion		
Gender	In favor	No opinion	Opposed
Male	330	165	55
Female	225	180	45

What proportion of the sampled females are in favor of more movies on campus?

a. 0.33
b. 0.5
c. 0.555
d. 0.6

ANSWER: b

125. A student organization is trying to decide whether or not to offer more movies on campus. The organization wants to determine whether this idea will appeal to members of both genders. A random sample of 1000 students were asked whether they were in favor of more movies on campus. The results by gender are shown in the table below.

	Opinion		
Gender	In favor	No opinion	Opposed
Male	330	165	55
Female	225	180	45

What proportion of the sampled males are in favor of more movies on campus?

a. 0.33
b. 0.5
c. 0.555

Chapter 2

d. 0.6

ANSWER: d

126. A student organization is trying to decide whether or not to offer more movies on campus. The organization wants to determine whether this idea will appeal to members of both genders. A random sample of 1000 students were asked if they were in favor of more movies on campus. The results by gender are shown in the table below.

	Opinion		
Gender	In favor	No opinion	Opposed
Male	330	165	55
Female	225	180	45

To answer the original question of whether or not to offer more movies on campus, which distribution should the student organization study?

a. The joint distribution of gender and opinion
b. The marginal distribution of gender
c. The conditional distribution of gender given opinion
d. The conditional distribution of opinion given gender

ANSWER: d

127. Prior to graduation, the members of a high school class were surveyed about their plans after high school. The table below displays the results by gender.

Gender	4-year college	2-year college	Military	Work	Other
Male	198	36	4	14	16
Female	176	36	1	3	5

If the data are going to be summarized by computing the marginal distribution of plans after high school, what should be the entry for "4-year college"?

a. 0.529
b. 0.739
c. 0.765
d. 374

ANSWER: c

128. Prior to graduation, the members of a high school class were surveyed about their plans after high school. The table below displays the results by gender.

Gender	4-year college	2-year college	Military	Work	Other
Male	198	36	4	14	16
Female	176	36	1	3	5

If the data are going to be summarized by computing the conditional distributions of plans after high school for male and female high school students, what should be the entry for "male" and "2-year college"?

a. 0.074
b. 0.134
c. 0.5

Chapter 2

d. 39.46

ANSWER: b

129. Prior to graduation, the members of a high school class were surveyed about their plans after high school. The table below displays the results by gender.

Gender	4-year college	2-year college	Military	Work	Other
Male	198	36	4	14	16
Female	176	36	1	3	5

If the data are going to be summarized by computing the conditional distributions of gender given plans after high school, what should be the entry for "male" and "2-year college"?

a. 0.074
b. 0.134
c. 0.36
d. 0.5

ANSWER: d

130. 130. A business has two types of employees: managers and workers. Managers earn either $100,000 or $200,000 per year. Workers earn either $10,000 or $20,000 per year. The numbers of male and female managers at each salary level and the numbers of male and female workers at each salary level are given in the table below.

Income	Male Managers	Female Managers
$100,000	80	20
$200,000	20	30
Income	Male Works	Female Workers
$10,000	30	20
$20,000	20	80

What is the proportion of male managers who make $200,000 per year?

a. 0.067
b. 0.133
c. 0.2
d. 0.4

ANSWER: c

131. A business has two types of employees: managers and workers. Managers earn either $100,000 or $200,000 per year. Workers earn either $10,000 or $20,000 per year. The numbers of male and female managers at each salary level and the numbers of male and female workers at each salary level are given in the table below.

Income	Male Managers	Female Managers
$100,000	80	20
$200,000	20	30
Income	Male Works	Female Workers
$10,000	30	20
$20,000	20	80

What is the proportion of female managers who make $200,000 per year?

Chapter 2

a. 0.1
b. 0.2
c. 0.4
d. 0.6

ANSWER: d

132. A business has two types of employees: managers and workers. Managers earn either $100,000 or $200,000 per year. Workers earn either $10,000 or $20,000 per year. The numbers of male and female managers at each salary level and the numbers of male and female workers at each salary level are given in the table below.

Income	Male Managers	Female Managers
$100,000	80	20
$200,000	20	30
Income	Male Works	Female Workers
$10,000	30	20
$20,000	20	80

What proportion of the managers are female?

a. 0.2
b. 0.333
c. 0.5
d. 0.667

ANSWER: b

133. A business has two types of employees: managers and workers. Managers earn either $100,000 or $200,000 per year. Workers earn either $10,000 or $20,000 per year. The numbers of male and female managers at each salary level and the numbers of male and female workers at each salary level are given in the table below.

Income	Male Managers	Female Managers
$100,000	80	20
$200,000	20	30
Income	Male Works	Female Workers
$10,000	30	20
$20,000	20	80

What conclusion(s) can we draw from this table?

a. The mean salary of female managers is greater than that of male managers.
b. The mean salary of males in this business is greater than the mean salary of females.
c. The mean salary of female workers is greater than that of male workers.
d. All of the above

ANSWER: d

134. A review of voter registration records in a small town yielded the following data for the numbers of males and females registered as Democrat, Republican, or some other affiliation.

Affiliation	Male	Female

Chapter 2

Democrat	300	600
Republican	500	300
Other	200	100

What proportion of male voters are registered Democrats?

a. 0.15
b. 0.30
c. 0.33
d. 300

ANSWER: b

135. A review of voter registration records in a small town yielded the following data for the numbers of males and females registered as Democrat, Republican, or some other affiliation.

Affiliation	Male	Female
Democrat	300	600
Republican	500	300
Other	200	100

What proportion of registered Democrats are male?

a. 0.15
b. 0.30
c. 0.33
d. 300

ANSWER: c

136. A review of voter registration records in a small town yielded the following data for the numbers of males and females registered as Democrat, Republican, or some other affiliation.

Affiliation	Male	Female
Democrat	300	600
Republican	500	300
Other	200	100

What proportion of all voters are male and registered as a Democrat?

a. 0.15
b. 0.30
c. 0.33
d. 300

ANSWER: a

137. Are avid readers more likely to wear glasses than those who read less frequently? Three-hundred men in Ohio were selected at random and characterized as to whether they wore glasses and whether the amount of reading they did was above average, average, or below average. The results are presented in the following table.

Amount of reading	Glasses? Yes	Glasses? No
Above average	47	26
Average	48	78

Chapter 2

Below average	31	70
Total	126	174

What is the proportion of men in the sample who wear glasses?

a. 0.24
b. 0.37
c. 0.42
d. 0.64

ANSWER: c

138. Are avid readers more likely to wear glasses than those who read less frequently? Three-hundred men in Ohio were selected at random and characterized as to whether they wore glasses and whether the amount of reading they did was above average, average, or below average. The results are presented in the following table.

Amount of reading	Glasses? Yes	Glasses? No
Above average	47	26
Average	48	78
Below average	31	70
Total	126	174

What is the proportion of all above-average readers who wear glasses?

a. 0.24
b. 0.37
c. 0.42
d. 0.64

ANSWER: d

139. A survey was conducted involving 303 subjects concerning their preferences with respect to the size of car they would consider purchasing. The following table shows the count of the responses by gender of the respondent.

Gender	Small car	Medium car	Large car	Total
Female	58	63	17	138
Male	79	61	25	165
Total	137	124	42	303

The data are to be summarized by constructing marginal distributions. In the marginal distribution for car size, the entry for medium cars is _____.

a. 0.457
b. 0.409
c. 0.370
d. 0.508
e. None of the above

ANSWER: b

140. A survey was conducted involving 303 subjects concerning their preferences with respect to the size of car they would consider purchasing. The following table shows the count of the responses by gender of the

Chapter 2

respondent.

Gender	Small car	Medium car	Large car	Total
Female	58	63	17	138
Male	79	61	25	165
Total	137	124	42	303

In the conditional distribution for preference of car size among male respondents, the entry for large cars is _____.

a. 0.056
b. 0.405
c. 0.152
d. 0.139
e. None of the above

ANSWER: c

141. A survey was conducted involving 303 subjects concerning their preferences with respect to the size of car they would consider purchasing. The following table shows the count of the responses by gender of the respondent.

Gender	Small car	Medium car	Large car	Total
Female	58	63	17	138
Male	79	61	25	165
Total	137	124	42	303

Among all respondents, the proportion of female respondents who preferred small cars is _____.

a. 0.420
b. 0.423
c. 0.452
d. 0.191
e. None of the above

ANSWER: d

142. Which of the following statements is(are) TRUE?

a. A two-way table is a useful way to summarize data when two categorical variables are measured on the same individuals or cases.
b. Simpson's paradox is an example of the potential effect of a lurking variable on an observed association between two categorical variables.
c. If the counts in each cell of a two-way table are divided by the total number of observations, the result is the joint distribution of the two categorical variables.
d. All of the above are true.
e. Only A and C are true.

ANSWER: d

143. A manufacturer is concerned about the quality of a particular item produced in one of its facilities. The facility operates with three 8-hour shifts of employees during the day. At the end of each shift, a random sample of the item is selected, and each item is checked against the required specifications. The selected items are

Chapter 2

classified as being either Acceptable or Unacceptable.
The following table summarizes the results,

	Shift		
Quality	One	Two	Three
Acceptable	96	118	64
Unacceptable	14	12	6
Total	110	130	70

Fill in the blanks in the following statement: Considering the entire day's production of all sampled items, the proportion produced by Shift One that are Unacceptable is _____. Among items produced by Shift One, the proportion of Unacceptable items is _____.

a. 0.127;0.045
b. 0.355;0.467
c. 0.045;0.127
d. 0.146;0.115
e. 0.103;0.045

ANSWER: c

144. A manufacturer is concerned about the quality of a particular item produced in one of its facilities. The facility operates with three 8-hour shifts of employees during the day. At the end of each shift, a random sample of the item is selected, and each item is checked against the required specifications. The selected items are classified as being either Acceptable or Unacceptable.
The following table summarizes the results:

	Shift		
Quality	One	Two	Three
Acceptable	96	118	64
Unacceptable	14	12	6
Total	110	130	70

The conditional distribution of Acceptable items produced by the three shifts is:

a. 0.355; 0.419; 0.226.
b. 0.345; 0.424; 0.230.
c. 0.873; 0.908; 0.914.
d. 0.310; 0.381; 0.206.
e. None of the above

ANSWER: b

145. Students at a small private university are required to take a basic statistics course. The table below summarizes one of the classes by gender and year in school.

	Year in School				
Gender	Freshman	Sophomore	Junior	Senior	TOTAL
Male	15	3	2	0	20
Female	50	20	8	2	80
TOTAL	65	23	10	2	100

What proportion of the students in the class are seniors?

Chapter 2

ANSWER: 0.02

146. Students at a small private university are required to take a basic statistics course. The table below summarizes one of the classes by gender and year in school.

	Year in School				
Gender	Freshman	Sophomore	Junior	Senior	TOTAL
Male	15	3	2	0	20
Female	50	20	8	2	80
TOTAL	65	23	10	2	100

What proportion of the students in the class are freshmen?

ANSWER: 0.65

147. Students at a small private university are required to take a basic statistics course. The table below summarizes one of the classes by gender and year in school.

	Year in School				
Gender	Freshman	Sophomore	Junior	Senior	TOTAL
Male	15	3	2	0	20
Female	50	20	8	2	80
TOTAL	65	23	10	2	100

What proportion of the students are male?

ANSWER: 0.20

148. Students at a small private university are required to take a basic statistics course. The table below summarizes one of the classes by gender and year in school.

	Year in School				
Gender	Freshman	Sophomore	Junior	Senior	TOTAL
Male	15	3	2	0	20
Female	50	20	8	2	80
TOTAL	65	23	10	2	100

What proportion of the students are female?

ANSWER: 0.80

149. If you have two categorical variables, one way to study them is to use a _____.

a. scatterplot
b. regression line
c. two-way table

ANSWER: c

150. 150. When possible, what is the best way to establish that an observed association is the result of a cause-and-effect relation?

a. Study the least-squares regression line.
b. Obtain the correlation coefficient.

Chapter 2

c. Examine *z*-scores rather than the original variables.
d. None of the above

ANSWER: d

151. 151. Which of the following is(are) necessary to establish a cause-and-effect relation between two variables?

a. A strong association between the variables
b. An association between the variables in many different settings
c. The alleged cause being plausible
d. All of the above

ANSWER: d

152. Let x = the midterm exam score and y = the final exam score for students in a large statistics class. The relationship between x and y is studied often. According to the instructor of this particular statistics class, those students who had an above-average value for x tended to have an above-average value for y. In other words, there was a positive association between x and y. Which of the following is(are) plausible explanations for this association?

a. Causation: x causes y. Thus, students who do well on the midterm exam should not worry about the final exam score, because the high score on the midterm will lead them to a high score on the final.
b. Common response: Changes in x and y are due to a common response to other variables. For example, students who spend much time studying will tend to do well on the midterm exam and on the final exam.
c. Common response: Changes in x and y are due to a common response to other variables. For example, students who do not understand statistics will tend to do poorly on the midterm exam and on the final exam. Likewise, students who understand statistics will tend to do well on the midterm exam and on the final exam.
d. Accidental: The association between x and y is purely coincidental. It is implausible to believe the observed association could be anything other than accidental.
e. Both B and C.

ANSWER: e

153. Let x = the number of people who failed to complete high school and y = the number of infant deaths. According to the 1990 census, those states having an above-average value for the variable x tended to have an above-average value for y. In other words, there was a positive association between x and y. What is the most plausible explanation for this association?

a. Causation: x causes y. Thus, programs to keep teens in school will help reduce the number of infant deaths.
b. Causation: y causes x. Thus, programs that reduce infant deaths will ultimately reduce the number of high school dropouts.
c. Common response: Changes in x and y are due to a common response to other variables. For example, states with large populations will have larger numbers of people who fail to complete high school and a larger number of infant deaths.
d. Accidental: The association between x and y is purely coincidental. It is implausible to believe the

observed association could be anything other than accidental.

ANSWER: c

154. Let x = the amount of money spent per pupil in high school. Let y = the mean Verbal SAT score for students taking the SAT. Recent data show that states having an above- average value for the variable x tend to have below-average values for the variable y. In other words, there is a negative association between x and y. This is particularly true in states having a large percent of all high school students taking the SAT. These states also tend to have larger populations. What is the most plausible explanation for this association?

a. Causation: x causes y. Overspending generally leads to extra, unnecessary programs, diverting attention from basic subjects. Inadequate training in these basic subjects generally leads to lower SAT scores.
b. Causation: y causes x. Low SAT scores create concerns about the quality of education. This inevitably leads to additional spending to help solve the problem.
c. Common response: Changes in x and y are due to a common response to other variables. If a higher percent of students take the SAT, the average score will be lower. Also, states with larger populations have large urban areas where the cost of living is higher and more money is needed for expenses.
d. Accidental: The association between x and y is purely coincidental. It is implausible to believe the observed association could be anything other than accidental.

ANSWER: c

155. As Swiss cheese matures, a variety of chemical processes take place. The taste of matured cheese is related to the concentration of several chemicals in the final product. Let x = the lactic acid concentration of the cheese and y = the taste score of the cheese, which was obtained by combining the scores from several tasters. According to a study of cheese in a certain region of Switzerland, those samples of cheese having an above-average value for the variable x tended to have an above-average value for y. In other words, there was a positive association between x and y. What is the most plausible explanation for this association?

a. Causation: x causes y. Thus, increasing the lactic acid concentration in cheese will generally help improve the taste.
b. Common response: Changes in x and y are due to a common response to other variables. For example, cheese samples that have matured longer will have higher lactic acid concentrations, and more mature cheese tends to receive higher taste scores.
c. Common response: Changes in x and y are due to a common response to other variables. For example, cheese with a higher concentration of hydrogen sulfide (another chemical present in cheese) will have higher lactic acid concentrations and will receive higher taste scores.
d. Accidental: Tthe association between x and y is purely coincidental. It is implausible to believe the observed association could be anything other than accidental.

ANSWER: a

156. An article in the student newspaper of a large university had the headline “A’s swapped for evaluations?” The article included the following paragraph:
According to a new study, teachers may be more inclined to give higher grades to students, hoping to gain favor with the university administrators who grant tenure. The study examined the average grade and teaching evaluation in a large number of courses in order to investigate the effects of grade inflation on evaluations. “I

Chapter 2

am concerned with student evaluations because instruction has become a popularity contest for some teachers," said Professor Smith, who recently completed the study. Results showed higher grades directly corresponded to a more positive evaluation.
Based on the statement underlined above, what did the study find out about the relationship between course grade and teaching evaluation?

a. Course grade is positively associated with teaching evaluation.
b. There must be a common response that course grade and teaching evaluation both respond to. Higher grades are usually obtained by the more serious students. These students are also more apt to fill out the course evaluation more seriously and positively.
c. The association between course grade and teaching evaluation is purely coincidental.
d. There is a cause-and-effect relationship between course grade and teaching evaluation.

ANSWER: a

157. An article in the student newspaper of a large university had the headline "A's swapped for evaluations?" The article included the following paragraph:
According to a new study, teachers may be more inclined to give higher grades to students, hoping to gain favor with the university administrators who grant tenure. The study examined the average grade and teaching evaluation in a large number of courses in order to investigate the effects of grade inflation on evaluations. "I am concerned with student evaluations because instruction has become a popularity contest for some teachers," said Professor Smith, who recently completed the study. Results showed higher grades directly corresponded to a more positive evaluation.
Which of the following would be a valid conclusion to draw from the study?

a. A teacher can improve his or her teaching evaluations by giving good grades.
b. A good teacher, as measured by teaching evaluations, helps students to learn better, resulting in higher grades.
c. Teachers of courses in which the mean grade is above average apparently tend to have above-average teaching evaluations.
d. All of the above

ANSWER: c

158. Which set of two variables is most likely to have a cause-and-effect relationship?

a. The height of a person and the weight of a person
b. The weight of a box and the postage rate one has to pay to ship the box to California
c. The make of a car and the mileage of the car
d. The age of a teacher and the income of the teacher

ANSWER: b

159. A researcher computed the average Math SAT score of all high school seniors who took the SAT exam for each of the 50 states. The researcher also computed the average salary of high school teachers in each of these states and plotted these average salaries against the average Math SAT scores for each state. The plot showed a distinct negative association between average Math SAT scores and average teacher salaries. The researcher can legitimately conclude which of the following?

a. Increasing the average salary of teachers will cause the average of Math SAT scores to decrease, but

it is not correct to conclude that increasing the salaries of individual teachers causes the Math SAT scores of individual students to increase.

b. States that pay teachers highly tend to do a poor job of teaching mathematics.

c. States whose students tend to perform poorly in mathematics probably have a higher proportion of problem students and thus need to pay teachers higher salaries in order to attract them to teach in those states.

d. The data used by the researcher do not provide evidence that increasing the salary of teachers will cause the performance of students on the Math SAT to get worse.

ANSWER: d

160. Fill in the blank. A researcher is conducting a study on contact lenses. She finds that, in a sample of contact-lens-wearing adults, those who wear their contact lenses for longer periods of time have more irritation in their eyes. Those who wear their contact lenses for only a few hours a day have less irritation. However, those who wear their contacts for longer times also tend to spend long hours at the office. As explanations for having more irritation in the eyes, the variables "length of time contacts are worn each day" and "length of time spent at the office per day" are _____ variables.

a. response
b. independent
c. confounding
d. placebo

ANSWER: c

161. Fill in the blank. A researcher notices that, in a sample of adults, those who take larger amounts of vitamin C have fewer illnesses. However, those who take larger amounts of vitamin C also tend to exercise more. As explanations for having fewer illnesses, the variables "amount of vitamin C taken" and "amount of exercise" are _____ variables.

a. skewed
b. confounding
c. response
d. symmetric

ANSWER: b

162. Which of the following statements about causation and association is FALSE?

a. Some observed associations between two variables are due to a lurking variable rather than to a cause-and-effect relationship between the two variables.

b. Some possible explanations of an observed association are causation, common response, and confounding.

c. When many variables interact with each other, confounding of several variables often prevents a conclusion about causation from being reached.

d. To establish that association is due to causation, it is best to conduct an experiment that makes changes to the explanatory variable, while controlling other influences on the response variable.

e. When strong association between two variables is present, this is often sufficient evidence to

establish that the association results from a causal link.

ANSWER: e

163. Two variables are confounded when their effects on a response variable can be distinguished from each other.

a. True
b. False

ANSWER: b

164. Interest rates for home mortgages have, in general, declined during recent months. With the apparent favorable influence for new-home building, there seems to be a clear relationship between x = the prevailing mortgage interest rate and y = the number of new houses being built per month in a Midwestern city over a period of 18 months. A scatterplot of the data collected shows that the linear model is appropriate. The equation of the least-squares regression line is
Number of new houses = 672.89 – 30.65 × Interest rate and $r^2 = 0.49$ Is the association between Interest rate and Number of new houses being built positive or negative?

ANSWER: Negative

165. Interest rates for home mortgages have, in general, declined during recent months. With the apparent favorable influence for new-home building, there seems to be a clear relationship between x = the prevailing mortgage interest rate and y = the number of new houses being built per month in a Midwestern city over a period of 18 months. A scatterplot of the data collected shows that the linear model is appropriate. The equation of the least-squares regression line is
Number of new houses = 672.89 – 30.65 × Interest rate and $r^2 = 0.49$ What is the correlation coefficient between Interest rate and Number of new houses being built?

ANSWER: –0.7

166. Interest rates for home mortgages have, in general, declined during recent months. With the apparent favorable influence for new-home building, there seems to be a clear relationship between x = the prevailing mortgage interest rate and y = the number of new houses being built per month in a Midwestern city over a period of 18 months. A scatterplot of the data collected shows that the linear model is appropriate. The equation of the least-squares regression line is
Number of new houses = 672.89 – 30.65 × Interest rate and $r^2 = 0.49$. Predict the Number of new houses being built for a month when the Interest rate is 10.2.

ANSWER: 360.26

167. Interest rates for home mortgages have, in general, declined during recent months. With the apparent favorable influence for new-home building, there seems to be a clear relationship between x = the prevailing mortgage interest rate and y = the number of new houses being built per month in a Midwestern city over a period of 18 months. A scatterplot of the data collected shows that the linear model is appropriate. The equation of the least-squares regression line is
Number of new houses = 672.89 – 30.65 × Interest rate and $r^2 = 0.49$
Which of the following descriptions best represents the value of the slope?

a. When no new houses are being built, the interest rate equals 30.65%.

Chapter 2

b. When the number of new houses being built increases by 1, the interest rate is expected to drop by 0.3065.

c. When the interest rate increases by 1%, the number of new houses being built is expected to drop by 30.65.

d. We cannot interpret the slope because we cannot build a negative number of new houses.

ANSWER: c

168. Interest rates for home mortgages have, in general, declined during recent months. With the apparent favorable influence for new-home building, there seems to be a clear relationship between x = the prevailing mortgage interest rate and y = the number of new houses being built per month in a Midwestern city over a period of 18 months. A scatterplot of the data collected shows that the linear model is appropriate. The equation of the least-squares regression line is

Number of new houses = 672.89 – 30.65 × Interest rate and $r^2 = 0.49$

What is the most plausible explanation for the association between Interest rate and Number of new houses being built?

a. Confounding

b. Common response

c. Causation

d. Accidental

ANSWER: b

169. A simple random sample of eight drivers was selected. All eight drivers are insured with the same insurance company, and all have similar auto insurance policies. The following table lists their driving experience (in years) and their monthly auto insurance premiums.

Driving experience (years)	5	2	12	9	15	6	25	16
Monthly auto insurance premium ($)	64	77	50	71	44	56	42	60

Which of the two variables is the explanatory variable?

ANSWER: Driving experience (years)

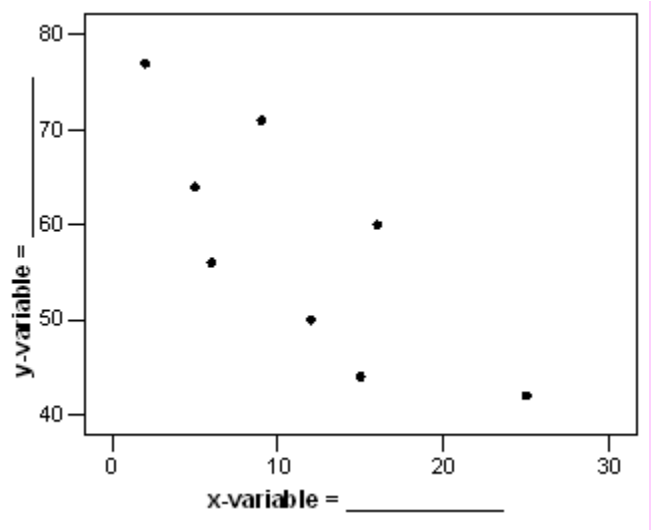
170. A simple random sample of eight drivers was selected. All eight drivers are insured with the same insurance company, and all have similar auto insurance policies. The following table lists their driving experience (in years) and their monthly auto insurance premiums.

Driving experience (years)	5	2	12	9	15	6	25	16
Monthly auto insurance premium ($)	64	77	50	71	44	56	42	60

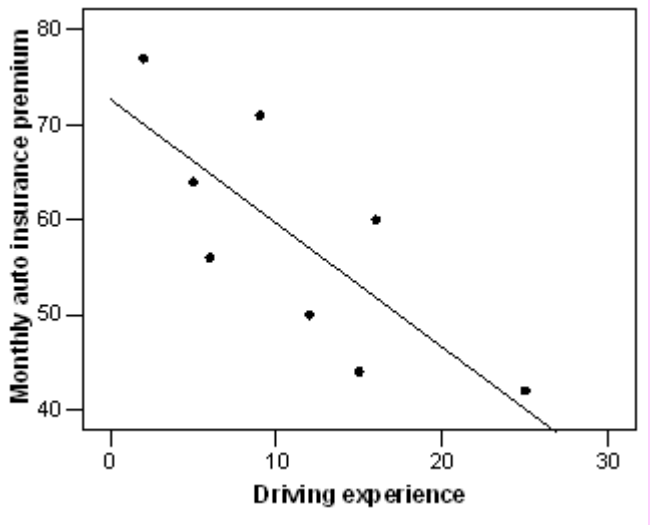
The equation of the least-squares regression line is $\hat{y} = 72.7 - 1.3x$. Sketch the line on the scatterplot below. Include the labels for the two axes.

Name: _____ Class: _____ Date: _____

Chapter 2



ANSWER:



171. A simple random sample of eight drivers was selected. All eight drivers are insured with the same insurance company, and all have similar auto insurance policies. The following table lists their driving experience (in years) and their monthly auto insurance premiums.

Driving experience (years)	5	2	12	9	15	6	25	16
Monthly auto insurance premium ($)	64	77	50	71	44	56	42	60

Complete the following sentence that explains what the value of the slope tells us:
When driving experience (choose one) increases / decreases by _____,
we expect the average monthly auto insurance premium to (choose one) increase / decrease by _____.

ANSWER: When driving experience *increases* by *1 year*, we expect the average monthly auto insurance premium to *decrease* by *$1.30*.

172. A simple random sample of eight drivers was selected. All eight drivers are insured with the same insurance company, and all have similar auto insurance policies. The following table lists their driving experience (in years) and their monthly auto insurance premiums.

Driving experience (years)	5	2	12	9	15	6	25	16
Monthly auto insurance	64	77	50	71	44	56	42	60

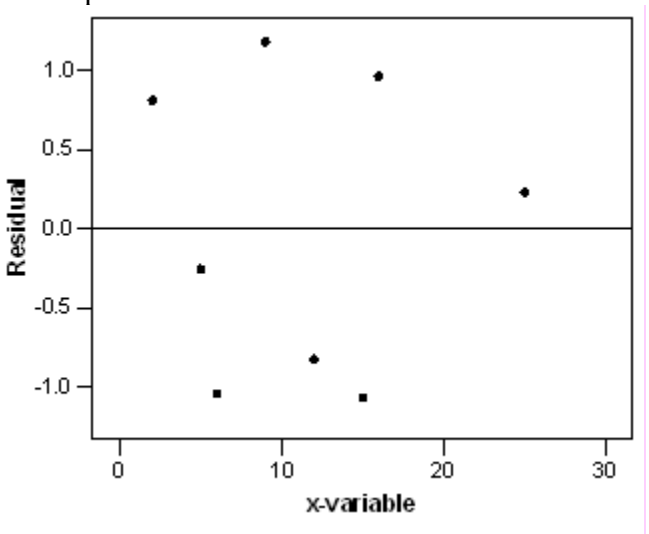
Chapter 2

premium ($)								

The correlation coefficient between driving experience and monthly auto insurance premium is $r = –0.775$. If we switched the roles of the explanatory and response variables, what would happen to the correlation coefficient?

ANSWER: The correlation coefficient would stay the same.

173. A plot of the residuals versus the x variable is shown.



Based on the residual plot, does the linear regression seem appropriate? Explain briefly.

ANSWER: Yes, because there is no clear pattern present in the graph.

174. Do heavier cars use more gasoline? To answer this question, a researcher randomly selected 15 cars. He collected data about the weight (in hundreds of pounds) and the mileage (mpg) for each car. From a scatterplot made with the data, a linear model seems appropriate.

The variable _____ is the response variable in this study.

ANSWER: mileage

175. Do heavier cars use more gasoline? To answer this question, a researcher randomly selected 15 cars. He collected data about the weight (in hundreds of pounds) and the mileage (mpg) for each car. From a scatterplot made with the data, a linear model seems appropriate.

The variable _____ is the explanatory variable in this study.

ANSWER: weight

176. Do heavier cars use more gasoline? To answer this question, a researcher randomly selected 15 cars. He collected data about the weight (in hundreds of pounds) and the mileage (mpg) for each car. From a scatterplot made with the data, a linear model seems appropriate.

The equation of the least-squares regression line is $\hat{y} = 40.4 – 0.521x$.

Which of the following descriptions of the value of the slope is the correct description?

a. The mileage is expected to decrease by 0.521 mpg when the weight of a car increases by 1 pound.

b. The mileage is expected to decrease by 0.521 mpg when the weight of a car increases by 100

Chapter 2

pounds.

c. The mileage is expected to decrease by 52.1 mpg when the weight of a car increases by 100 pounds.

d. We cannot interpret the slope because we cannot have a negative weight of a car.

ANSWER: b

177. Do heavier cars use more gasoline? To answer this question, a researcher randomly selected 15 cars. He collected data about the weight (in hundreds of pounds) and the mileage (mpg) for each car. From a scatterplot made with the data, a linear model seems appropriate.
The percent of variation in mileage that is accounted for by the linear relationship between mileage and weight is approximately 44%. What is the value of the correlation coefficient between the weight and the mileage of a car?

ANSWER: $r = -0.663$