Multiple Choice
Identify the choice that best completes the statement or answers the question.

1. An office supply catalog gives a description of bookshelves that includes the following variables. Which of these variables is categorical?
   A. The width of a bookshelf in inches
   B. The width of a bookshelf in feet
   C. The color of the bookshelf
   D. The height of a bookshelf in inches
   E. The weight of a bookshelf

2. A professor records the values of several variables for each student in her class. These include the variables listed below. Which of these variables is categorical?
   A. Score on the final exam (out of 200 points).
   B. Final grade for the course (A, B, C, D, or F).
   C. The total number of points earned in the class (i.e., the total of the points on all exams and quizzes in the course; the maximum number of points possible is 500).
   D. The number of lectures the student missed.
   E. Amount of time, in minutes, spent studying for the final exam.

3. A survey typically records many variables of interest to the researchers involved. Below are some of the variables from a survey conducted by the U.S. Postal Service. Which of the variables is categorical?
   A. County of residence
   B. Number of people, both adults and children, living in the household
   C. Total household income, before taxes, in 1993
   D. Age of respondent
   E. Number of rooms in the dwelling

4. A particularly common question in the study of wildlife behavior involves observing contests between "residents" of a particular area and "intruders." In each contest, the "residents" either win or lose the encounter (assuming there are no ties). Observers might record several variables, listed below. Which of these variables is categorical?
   A. The duration of the contest (in seconds).
   B. The number of animals involved in the contest.
   C. Whether the "residents" win or lose.
   D. The total number of contests won by the "residents."
   E. None of these.

5. You measure the age, marital status and earned income of an SRS of 1463 women. The number and type of variables you have measured is
   A. 14563.
   B. four; two categorical and two quantitative.
   C. four; one categorical and three quantitative.
   D. three; two categorical and one quantitative.
   E. three; one categorical and two quantitative.

6. You measure the age (years), weight (pounds), and marital status (single, married, divorced, or widowed) of 1400 women. How many variables did you measure?
   A. 1
7. A statistics teacher asks the 29 students in his statistics class how many minutes they spent on one homework assignment. The distribution of the variable “time on homework” is
A. the difference between the longest time and the shortest time among the students’ responses.
B. a description of what values the variable takes and how often it takes them.
C. the average distance between each value of the variable.
D. the average time the students spent on the assignment.
E. the number of students who were asked the questions—that is, 29.

8. You open a package of plain M & M candies and count how many there are of each color. The distribution of the variable “candy color” is:
A. The colors: Red, Orange, Green, Yellow, Brown, and Blue.
B. The total number of candies in the package.
C. Six—the number of different colors there are in the package.
D. The six different colors and how many there are of each.
E. Since “color” is a categorical variable, it doesn’t have a distribution.

9. Deciduous forests in the Eastern United States often have many different species of oak trees. Below is a frequency distribution for five different species of oaks found in sample plots of a certain forest.

<table>
<thead>
<tr>
<th>Species of oak</th>
<th>Black</th>
<th>Red</th>
<th>Scarlet</th>
<th>Pin</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>25</td>
<td>14</td>
<td>12</td>
<td>8</td>
<td>40</td>
</tr>
</tbody>
</table>

Which of the following pie charts describes the same distribution?

A.  
B.  
D.  
E.  

10. X and Y are two categorical variables. The best way to determine if there is a relation between them is to
   A. construct parallel box plots of the X and Y values.
   B. draw dot plots of the X and Y values.
   C. make a two-way table of the X and Y values.
   D. compare medians and interquartile ranges of the X and Y values.
   E. compare means and standard deviations of the X and Y values.

11. In a study of the link between high blood pressure and cardiovascular disease, a group of white males aged 35 to 64 was followed for 5 years. At the beginning of the study, each man had his blood pressure measured and it was classified as either "low" systolic blood pressure (less than 140 mm Hg) or "high" blood pressure (140 mm Hg or higher). The following table gives the number of men in each blood pressure category and the number of deaths from cardiovascular disease during the 5-year period.

<table>
<thead>
<tr>
<th>Blood pressure</th>
<th>Deaths</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>10</td>
<td>2000</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>3500</td>
</tr>
</tbody>
</table>

Based on these data, which of the following statements is correct?
   A. These data are consistent with the idea that there is a link between high blood pressure and death from cardiovascular disease.
   B. The mortality rate (proportion of deaths) for men with high blood pressure is 5 times that of men with low blood pressure.
   C. These data probably understate the link between high blood pressure and death from cardiovascular disease, because men will tend to understate their true blood pressure.
   D. Although there were more deaths in the high blood pressure group, this is expected, because there were 1500 more men in that group.
   E. All of the above.

Scenario 1-1
A review of voter registration records in a small town yielded the following table of the number of males and females registered as Democrat, Republican, or some other affiliation.

<table>
<thead>
<tr>
<th>Affiliation</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrat</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>Republican</td>
<td>500</td>
<td>300</td>
</tr>
<tr>
<td>Other</td>
<td>200</td>
<td>100</td>
</tr>
</tbody>
</table>

12. Use Scenario 1-1. The proportion of males that are registered as Democrats is
   A. 300
   B. 30
13. Use Scenario 1-1. Your percentage from question number 12 is part of
   A. The marginal distribution of political party registration.
   B. The marginal distribution of gender.
   C. The conditional distribution of gender among Democrats.
   D. The conditional distribution of political party registration among males.
   E. The conditional distribution of males within gender.

14. Use Scenario 1-1. The proportion of registered Democrats that are male is
   A. 300
   B. 33
   C. 0.33
   D. 0.30
   E. 0.15

15. Use Scenario 1-1. Your percentage from question number 14 is part of
   A. The marginal distribution of political party registration.
   B. The marginal distribution of gender.
   C. The conditional distribution of gender among Democrats.
   D. The conditional distribution of political party registration among males.
   E. The conditional distribution of males within gender.

16. Use Scenario 1-1. Which of the following graphs accurately represents the distribution for political party registration for each gender?
   A. 
   B. 
   C. 
   D. 
   E. 

   ![Graph A](image1)
   ![Graph B](image2)
   ![Graph C](image3)
   ![Graph D](image4)
   ![Graph E](image5)
### Scenario 1-2

Below is a two-way table summarizing the number of cylinders in selected car models manufactured in six different countries in the 1990’s.

<table>
<thead>
<tr>
<th>Number of cylinders</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
</tr>
</tbody>
</table>

**Table:**

<table>
<thead>
<tr>
<th>Country</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Italy</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Japan</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Sweden</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>8</td>
<td>22</td>
</tr>
</tbody>
</table>

**Questions:**

17. Use Scenario 1-2. The percentage of all cars listed in the table with 4-cylinder engines is
   A. 19%.
   B. 21%.
   C. 50%.
   D. 80%.
   E. 91%.

18. Use Scenario 1-2. The percent of cars with 4-cylinder engines that are made in Germany is
   A. 10.5%.
   B. 21%.
   C. 50%.
   D. 80%.
   E. 91%.

19. Use Scenario 1-2. Which of the following is a marginal distribution?
   A. The percentage of all four-cylinder cars manufactured in Germany.
   B. The number of four-cylinder cars manufactured in Germany.
   C. The percentage of all cars manufactured in each country.
   D. The percentage of cars manufactured in Germany for each number of cylinders.
   E. The numbers 4, 5, 6, 8.

20. Use Scenario 1-2. From this table, we might conclude that
   A. there is a strong association between country of origin and number of cylinders.
   B. about 18% of the cars sold in the United States were manufactured in Japan.
C. these data could be more effectively presented with a box plot.
D. the only eight cylinder cars in this data set were manufactured in Germany.
E. All the cars on Italian roads have four cylinders.

21. The bar graph below summarizes responses of dog owners to the question, “Where in the car do you let your dog ride?”

Which of the following statements is false?
A. Some owners let their pets ride in more than one place in the car.
B. A majority of owners allow their pets to ride in the front passenger seat.
C. The most common place dogs ride is in the back seat.
D. The vertical scale of this graph exaggerates the difference between the percentage who let their dogs ride in the driver’s lap versus a passenger’s lap.
E. These data could also be presented in a pie chart.

22. The table below shows the results of the New Hampshire Democratic Presidential Primary on January 8, 2008.

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Percentage of votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hillary Clinton</td>
<td>39</td>
</tr>
<tr>
<td>Barack Obama</td>
<td>37</td>
</tr>
<tr>
<td>John Edwards</td>
<td>17</td>
</tr>
<tr>
<td>Bill Richardson</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
</tr>
</tbody>
</table>

Which of the following lists of graphs are all appropriate ways of presenting these data?
A. Bar graph, Pie Chart, Box plot
B. Bar graph, Box plot
C. Bar graph, Pie Chart
D. Bar Graph only
E. Pie Chart only

23. One way economists measure the health of the real estate market is by counting “housing starts,” or the number of permits issued for construction of new homes. Below is a graph displaying housing starts (in thousands) in the United States from 2006 to 2009.
What is the principle weakness of this graphical presentation of data?
A. The “thousands” label on the vertical scale is confusing and misleading.
B. The data only shows housing starts for four years, which is not enough time to identify a meaningful trend.
C. Using proportionally-sized pictograms exaggerates the difference between years.
D. Data of this type should only be displayed in a pie chart.
E. It is unclear which dimension of the house represents the number of housing starts for that year.

24. The histogram below show the length (in minutes) of 140 songs recorded by the band Wilco.

Which of the following descriptions best fits this distribution?
A. Skewed right, centered at about 8, with several high outliers.
B. Skewed left, centered at about 8, with several high outliers.
C. Skewed right, centered at about 4.5, with several high outliers.
D. Skewed left, centered at about 4.5, with several high outliers.
E. Skewed left, centered at about 3.5, with several high outliers.

Figure 1-1
25. Use Figure 1-1. For these data,
   A. the median jump is between 75 and 80 inches.
   B. the median jump is between 80 and 85 inches.
   C. the smallest jump must be below 65 inches.
   D. the winning jump in the 1976 Olympic Games was 40 inches.
   E. none of the above.

26. Use Figure 1-1. The mean of this histogram is approximately
   A. 70 inches.
   B. 74 inches.
   C. 78 inches.
   D. 82 inches.
   E. 86 inches.

27. Use Figure 1-1. Based on this histogram, the percentage of the winning jumps that were at least 80 inches is about
   A. 10%.
   B. 35%.
   C. 45%.
   D. 55%.
   E. 90%.

28. The following histogram represents the distribution of acceptance rates (percent accepted) among 25 business schools in 1997.
29. The histogram below shows the distribution of heights for 100 randomly selected school children in Great Britain.

Which of the following descriptions best fits this distribution?
A. Roughly uniform, centered at about 150, range 110 to 190.
B. Roughly uniform, centered at about 150, range 80.
C. Roughly symmetric, centered at about 150, range 110 to 190.
D. Roughly symmetric, centered at about 150, range 80.
E. Roughly symmetric, centered at about 150, range about 135 to 165.

Scenario 1-3
For a physics course containing 10 students, the maximum point total for the quarter was 200. The point totals for the 10 students are given in the stemplot below.
30. Use Scenario 1-3. This stemplot is most similar to
A. a histogram with class intervals $110 \leq \text{score} < 120$, $120 \leq \text{score} < 130$, etc.
B. a time plot of the data with the observations taken in increasing order.
C. a boxplot of the data.
D. reporting the 5 number summary for the data, with the mean.
E. a dot plot of the data.

31. Use Scenario 1-3. To which of the following data sets does this stemplot correspond?
A. All integers between 116 and 179
B. 1, 2, 3, 4, 6, 6, 7, 8, 8, 9
C. 16, 18, 21, 24, 28, 33, 37, 42, 46, 79
D. 116, 118, 121, 124, 128, 133, 137, 142, 146, 179
E. None of the above.

32. Use Scenario 1-3. The median point total for this class is
A. 130.
B. 130.5.
C. 133.
D. 134.4.
E. 137.

33. Which of the following statements is NOT true?
A. In a symmetric distribution, the mean and the median are equal.
B. Fifty percent of the scores in a distribution are between the first and third quartiles.
C. In a symmetric distribution, the median is halfway between the first and third quartiles.
D. The median is always greater than the mean.
E. The range is the difference between the largest and the smallest observation in the data set.

34. When drawing a histogram it is important to
A. have a separate class interval for each observation to get the most informative plot.
B. make sure the heights of the bars exceed the widths of the class intervals so that the bars are true rectangles.
C. label the vertical axis so the reader can determine the counts or percent in each class interval.
D. leave large gaps between bars. This allows room for comments.
E. scale the vertical axis according to the variable whose distribution you are displaying.

35. A consumer group surveyed the prices for a certain item in five different stores, and reported the average price as $15. We visited four of the five stores, and found the prices to be $10, $15, $15, and $25. Assuming that the consumer group is correct, what is the price of the item at the store that we did not visit?
A. $5
B. $10
36. The ages of people in a college class are as follows:

<table>
<thead>
<tr>
<th>Age</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>19</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
</tr>
<tr>
<td>21</td>
<td>200</td>
</tr>
<tr>
<td>22</td>
<td>90</td>
</tr>
<tr>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
</tr>
</tbody>
</table>

What is true about the median age?
A. It must be 20.
B. It must be 20.5.
C. It could be any number between 19 and 21.
D. It must be 21.
E. It must be over 21.

37. The median age of five elephants at a certain zoo is 30 years. One of the elephants, whose age is 50 years, is transferred to a different zoo. The median age of the remaining four elephants is
A. 40 years.
B. 30 years.
C. 25 years.
D. less than 30 years.
E. Cannot be determined from the information given.

38. A set of data has a mean that is much larger than the median. Which of the following statements is most consistent with this information?
A. The distribution is symmetric.
B. The distribution is skewed left.
C. The distribution is skewed right.
D. The distribution is bimodal.
E. The data set probably has a few low outliers.

39. A random sample of 100 students in grades 10 through 12 were sampled and asked their year in school and whether they were involved in interscholastic sports, intramural sports, or no sports. The results are summarized in the segmented bar graph below.

Based on this graph, which of the following statements is true?
A. More seniors are involved in interscholastic sports than sophomores.
B. There is no association between year in school and whether students are involved in sports.
C. There were more seniors in the sample than juniors.
D. Juniors have the highest percentage participation in intramurals.
E. Less than half the seniors are involved in either interscholastic or intramural sports.

40. Which of the following graphs can be used to summarize the data in a two-way table?
A. Dot plot
B. Segmented bar graph
C. Box plot
D. Stem plot
E. Histogram

Scenario 1
Mr. Williams asked the 26 seniors in his statistics class how many A.P. courses they had taken during high school. Below is a dot plot summarizing the results of his survey.

41. Use Scenario 1. The median number of A.P. courses taken by Mr. Williams’s students is
A. 2
B. 3
C. 3.5
D. 4
E. cannot be determined from the information given.

42. Use Scenario 1. The interquartile range for the number of A.P. Courses is
A. 3 to 4
B. 2.5 to 5
C. 3 to 5
D. 2
E. 2.5

43. Use Scenario 1. Which of the following is a correct box plot for these data?
44. The mean age of four people in a room is 30 years. A new person whose age is 55 years enters the room. The mean age of the five people now in the room is
A. 30.
B. 35.
C. 37.5.
D. 40.
E. Cannot be determined from the information given.

Scenario 1-5
A sample was taken of the salaries of 20 employees of a large company. The following boxplot shows the salaries (in thousands of dollars) for this year.

45. Use Scenario 1-5. Based on the boxplot, which of the following statements is true?
A. The maximum salary is between $60,000 and $70,000.
B. The minimum salary is $20,000.
C. The range of the middle half of the salaries is about $20,000.
D. The median salary is about $40,000.
E. 25% of the employees make more than $70,000.

46. Use Scenario 1-5. Based on the boxplot, the five-number summary is
A. 28, 39, 48, 60.5, 77.
B. 28, 41, 48, 58, 77.
C. 28, 39, 51, 58, 77.
D. 28, 41, 51, 60.5, 77.
E. 26, 39, 48, 60.5, 81.

Scenario 1-6
The following is a boxplot of the birth weights (in ounces) of a sample of 160 infants born in a local hospital.

47. Use Scenario 1-6. The median birthweight is approximately
A. 80.5 ounces.
B. 90 ounces.
C. 100 ounces.
D. 110 ounces.
E. 120 ounces.

48. Use Scenario 1-6. About 40 of the birthweights were below
A. 92 ounces.
B. 102 ounces.
C. 112 ounces.
D. 122 ounces.
E. 132 ounces.

49. Use Scenario 1-6. The number of children with birthweights between 102 and 122 ounces is approximately:
A. 20.
B. 40.
C. 50.
D. 80.
E. 100.

50. There are three children in a room, ages three, four, and five. If a four-year-old child enters the room the
A. mean age will stay the same but the variance will increase.
B. mean age will stay the same but the variance will decrease.
C. mean age and variance will stay the same.
D. mean age and variance will increase.
E. mean age and variance will decrease.
51. The standard deviation of 16 peoples’ weights (in pounds) is computed to be 5.4. The variance of these measurements is
   A. 2.24.
   B. 29.16.
   C. 52.34.
   D. 256.
   E. 21.6.

52. The standard deviation of 16 peoples’ weights (in pounds) is computed to be 5.4. The units for the variance of these measurements is
   A. pounds.
   B. square root pounds.
   C. pounds squared.
   D. no units. Variance never has units.
   E. percentiles.

53. A sample of production records for an automobile manufacturer shows the following figures for production per shift:

   705 700 690 705

   The variance of the sample is
   A. 8.66.
   B. 7.07.
   C. 75.00.
   D. 50.00.
   E. 20.00.

54. You catch 10 cockroaches in your bedroom and measure their lengths in centimeters. Which of these sets of numerical descriptions are all measured in centimeters?
   A. median length, variance of lengths, largest length
   B. median length, first and third quartiles of lengths
   C. mean length, standard deviation of lengths, median length
   D. mean length, median length, variance of lengths.
   E. both (B) and (C)

55. A policeman records the speeds of cars on a certain section of roadway with a radar gun. The histogram below shows the distribution of speeds for 251 cars.
Which of the following measures of center and spread would be the best ones to use when summarizing these data?
A. Mean and interquartile range.
B. Mean and standard deviation.
C. Median and range.
D. Median and standard deviation.
E. Median and interquartile range.

56. You want to use numerical summaries to describe a distribution that is strongly skewed to the left. Which combination of measure of center and spread would be the best ones to use?
A. Mean and interquartile range.
B. Mean and standard deviation.
C. Median and range.
D. Median and standard deviation.
E. Median and interquartile range.

57. A lobster fisherman is keeping track of the productivity of a set of traps he has placed in a favorite location. Below are the numbers of lobsters in these traps over the course of 12 different hauls.
0 3 3 3 4 5 5 6 7 7 12 14

According to the 1.5 x IQR rule, which values in the above distribution are outliers?
A. 0 only
B. 14 only
C. 12 and 14
D. 0 and 14
E. 0, 12, and 14

58. The first sentence in Henry James’s novel *The Turn of the Screw* has 62 words. The five number summary for the lengths of those words is 1, 2, 3, 5, 6, 12. According to the 1.5 x IQR rule for identifying outliers, does this distribution have any outliers?
A. No, there are no outliers.
B. Yes, there is at least one high outlier but no low outliers.
C. Yes, there is at least one low outlier, but no high outliers.
D. Yes, there is at least one high and one low outlier.
E. There is not enough information given to determine if there are any outliers.

59. Different writers have different styles. One way to quantify this difference is to compare the distribution of word lengths in their work. Below are parallel boxplots describing the distributions of word lengths for the first 60 words in Henry James’s *The Turn of the Screw*, J.K. Rowling’s *Harry Potter and the Chamber of Secrets*, and Chapter 1 of your statistics textbook (labeled “Starnes” below).
Based on the graphs, which one of the following statements must be true?
A. Dot plots of the distributions of James’s word lengths and Starnes’s word lengths are identical.
B. The longest word in the distribution of Rowling’s word lengths is shorter than 25% of the word in the “James” distribution.
C. The range of Rowling’s word lengths is smaller than the interquartile range of Starnes’s word lengths.
D. The median word length for Rowling is longer than for either Starnes or James.
E. 75% of the words in Rowling’s distribution are longer than the median word length in Starnes’s distribution.

60. The stemplot below shows the number of home runs hit in 2008 by members of the Philadelphia Phillies, who won major League Baseball’s World Series that year. (Each of the 13 players who appeared in at least half the Phillies’ games that year is included). Note that 4 8 represents 48 home runs.

```
0 | 0 2 4
0 | 9 9 9
1 | 1 4 4
1 |
2 | 4
2 |
3 | 3 3
3 |
4 |
4 | 8
```

The five number summary for these data is:
A. 0, 9, 1, 3, 8
B. 0, 9, 11, 33, 48
C. 0, 6.5, 11, 28.5, 48
D. 0, 6.5, 11, 28.5, 33
E. 0, 4, 11, 24, 48
61. A study of the salaries of full professors at Upper Wabash Tech shows that the median salary for female professors is considerably less than the median male salary. 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. This apparent contradiction is an example of
A. a fallacy.
B. Simpson's paradox.
C. concealed gender bias.
D. a conditional distribution.
E. negative association.

62. The reversal of the direction of an association when a lurking variable is taken into account is called
A. a conditional association.
B. lurking association.
C. Simpson's paradox.
D. negative association.
E. independence.

63. Alexa's school newspaper publishes an article saying that a poll of 200 male and female students indicated that 60% of the male students did the summer reading, while only 45% of the female students did the summer reading. Alexa suspects that this is a distortion of the true facts, and that Simpson's paradox is to blame. She suspects that grade level (9th through 12th) is a lurking variable. What should she do to investigate her suspicions?
A. Draw parallel box plots of the summer reading data for males and females to see if there is a difference in the shape or center of the two distributions.
B. Undertake a new poll and only ask students in grades 12 about the summer reading.
C. Look at the conditional distributions in a two-way table of gender versus yes/no for summer reading.
D. Make two two-way tables, one for males, one for females, in which the two variables are grade level and yes/no for summer reading.
E. Compare grade level and yes/no for summer reading in one two-way table, without dividing students according to gender.

Scenario 3-1
The height (in feet) and volume (in cubic feet) of usable lumber of 32 cherry trees are measured by a researcher. The goal is to determine if volume of usable lumber can be estimated from the height of a tree.

64. Use Scenario 3-1. In this study, the response variable is
A. height of researcher.
B. volume of lumber.
C. height of tree.
D. the measuring instrument used to measure volume.
E. impossible to determine.

65. Use Scenario 3-1. Which of the following statements are supported by the scatterplot?
I. There is a positive association between height and volume.
II. There is an outlier in the plot.
III. As the height of a cherry tree increases, the volume of useable lumber it yields increases.
A. I only
B. II only
C. III only
D. I and II
E. I, II, and III

66. Use Scenario 3-1. If the data point (65,70) were removed from this study, how would the value of the correlation $r$ change?
A. $r$ would be smaller, since there are fewer data points.
B. $r$ would be smaller, because this point falls in the pattern of the rest of the data.
C. $r$ would be larger, since the $x$ and $y$ coordinates are larger than the mean $x$ and mean $y$, respectively.
D. $r$ would be larger, since this point does not fall in the pattern of the rest of the data.
E. $r$ would not change, since it’s value does not depend which variable is used for $x$ and which is used for $y$.

67. A study is conducted to determine if one can predict the yield of a crop based on the amount of fertilizer applied to the soil. The response variable in this study is
A. yield of the crop.
B. amount of fertilizer applied to the soil.
C. the experimenter.
D. amount of rainfall.
E. the soil.

68. 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). In this study, the explanatory variable is
A. amount of eroded soil.
B. rate of water flow.
C. size of soil bed.
D. depth of soil bed.
E. liters/second.

69. Two variables are said to be negatively associated if
A. larger values of one variable are associated with larger values of the other.
B. larger values of one variable are associated with smaller values of the other.
C. smaller values of one variable are associated with smaller values of the other.
D. smaller values of one variable are associated with both larger or smaller values of the other.
E. there is no pattern in the relationship between the two variables.

70. You would draw a scatterplot to
A. show the distribution of heights of students in this course.
B. compare the distributions of heights for male and female students in this course.
C. show the relationship between gender and having a driver’s license.
D. show the five-number summary for the heights of female students.
E. show the relationship between the height of female students and the heights of their mothers.

71. A study of the effects of television on child development measured how many hours of television each of 125 grade school children watched per week during a school year and each child’s reading score. Which variable would you put on the horizontal axis of a scatterplot of the data?
A. Reading score, because it is the response variable.
B. Reading score, because it is the explanatory variable.
C. Hours of television, because it is the response variable.
D. Hours of television, because it is the explanatory variable.
E. It makes no difference, because there is no explanatory-response distinction in this study.

Scenario 3-2
The following table and scatter plot present data on wine consumption (in liters per person per year) and death rate from heart attacks (in deaths per 100,000 people per year) in 19 developed Western countries.

### Wine Consumption and Heart Attacks

<table>
<thead>
<tr>
<th>Country</th>
<th>Alcohol from wine</th>
<th>Heart disease deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>2.5</td>
<td>211</td>
</tr>
<tr>
<td>Austria</td>
<td>3.9</td>
<td>167</td>
</tr>
<tr>
<td>Belgium</td>
<td>2.9</td>
<td>131</td>
</tr>
<tr>
<td>Canada</td>
<td>2.4</td>
<td>191</td>
</tr>
<tr>
<td>Denmark</td>
<td>2.9</td>
<td>220</td>
</tr>
<tr>
<td>Finland</td>
<td>0.8</td>
<td>297</td>
</tr>
<tr>
<td>France</td>
<td>9.1</td>
<td>71</td>
</tr>
<tr>
<td>Iceland</td>
<td>0.8</td>
<td>211</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.7</td>
<td>300</td>
</tr>
<tr>
<td>Italy</td>
<td>7.9</td>
<td>107</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.8</td>
<td>167</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1.9</td>
<td>266</td>
</tr>
<tr>
<td>Norway</td>
<td>0.8</td>
<td>227</td>
</tr>
<tr>
<td>Spain</td>
<td>6.5</td>
<td>86</td>
</tr>
<tr>
<td>Sweden</td>
<td>1.6</td>
<td>115</td>
</tr>
<tr>
<td>Switzerland</td>
<td>5.8</td>
<td>285</td>
</tr>
<tr>
<td>United</td>
<td>1.3</td>
<td>199</td>
</tr>
<tr>
<td>Kingdom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>1.2</td>
<td>172</td>
</tr>
<tr>
<td>West Germany</td>
<td>2.7</td>
<td></td>
</tr>
</tbody>
</table>

![Scatter plot: Heart disease deaths vs alcohol from wine](image)
72. Use Scenario 3-2. The scatterplot shows that
A. countries that drink more wine have higher death rates from heart disease.
B. the amount of wine a country drinks is not related to its heart disease death rate.
C. countries that drink more wine have lower death rates from heart disease.
D. heart disease deaths is the explanatory variable.
E. country is the explanatory variable.

73. Use Scenario 3-2. Which country is represented by the clear triangle in the scatter plot?
A. New Zealand
B. Canada
C. Finland
D. Belgium
E. Italy

74. Use Scenario 3-2. Do these data provide strong evidence that drinking wine actually *causes* a reduction in heart disease deaths?
A. Yes. The strong straight-line association in the plot shows that wine has a strong effect on heart disease deaths.
B. No. Countries that drink lots of wine may differ in other ways from countries that drink little wine. We can't be sure the wine accounts for the difference in heart disease deaths.
C. No. \( r \) does not equal \(-1\).
D. No. The plot shows that differences among countries are not large enough to be important.
E. No. The plot shows that deaths go up as more alcohol from wine is consumed.

75. Use Scenario 3-2. The correlation between wine consumption and heart disease deaths is one of the following values. From the scatterplot, which must it be?
A. \( r = -0.84 \)
B. \( r = -0.25 \)
C. \( r \) is very close to 0
D. \( r = 0.25 \)
E. \( r = 0.84 \)

76. Use Scenario 3-2. If heart disease death rate were expressed as deaths per 1,000 people instead of as deaths per 100,000 people, how would the correlation \( r \) between wine consumption and heart disease death rate change?
A. \( r \) would be divided by 100.
B. \( r \) would be divided by 10.
C. \( r \) would not change.
D. \( r \) would be multiplied by 10.
E. \( r \) would be multiplied by 100.

77. Use Scenario 3-2. The wine consumption data are in liters of alcohol per person. Which of these are *all* measured in liters of alcohol per person?
A. The mean, the first quartile, and the variance of wine consumption.
B. The median wine consumption and the correlation between wine consumption and heart disease deaths.
C. The median, the variance, and the standard deviation of wine consumption.
D. The standard deviation of wine consumption and the correlation between wine consumption and heart disease deaths.
E. The mean, the median, and the standard deviation of wine consumption.
78. There is a positive correlation between the size of a hospital (measured by number of beds) and the median number of days that patients remain in the hospital. Does this mean that you can shorten a hospital stay by choosing to go to a small hospital?

A. No – a negative correlation would allow that conclusion, but this correlation is positive.
B. Yes – the data show that stays are shorter in smaller hospitals.
C. No – the positive correlation is probably explained by the fact that seriously ill people go to large hospitals.
D. Yes – the correlation can't just be an accident.
E. Yes – but only if r is very close to 1.

79. Below is a scatterplot of wine consumption (in liters per person per year) and death rate from heart attacks (in deaths per 100,000 people per year) in 19 developed Western countries. European countries are designated by closed circles, other countries are designated by open circles.

![Heart disease deaths vs alcohol from wine](scatterplot.png)

Which of the following statement is not supported by the information in the scatter plot?

A. About half the European countries consume more wine per person than any of the non-European countries.
B. On average, the non-European countries drink less wine and have more heart attacks.
C. The four countries with the highest rates of wine consumption are all European.
D. The correlation between wine consumption and heart disease deaths is equally strong in European countries and non-European countries.
E. The country with the highest heart disease death rate is in Europe.

80. The correlation coefficient measures

A. whether there is a relationship between two variables.
B. the strength of the relationship between two quantitative variables.
C. whether or not a scatterplot shows an interesting pattern.
D. whether a cause and effect relation exists between two variables.
E. the strength of the linear relationship between two quantitative variables.

81. Which of the following are most likely to be negatively correlated?

A. The total floor space and the price of an apartment in New York.
B. The percentage of body fat and the time it takes to run a mile for male college students.
C. The heights and yearly earnings of 35-year-old U.S. adults.
D. Gender and yearly earnings among 35-year-old U.S. adults.
E. The prices and the weights of all racing bicycles sold last year in Chicago.

Scenario 3-3
Consider the following scatterplot, which describes the relationship between stopping distance (in feet) and air temperature (in degrees Centigrade) for a certain 2,000-pound car travelling 40 mph.

82. Use Scenario 3-3. The correlation between temperature and stopping distance
   A. is approximately 0.9.
   B. is approximately 0.6.
   C. is approximately 0.0.
   D. is approximately -0.6.
   E. cannot be calculated, because some of the x values are negative.

83. Use Scenario 3-3. If another data point were added with an air temperature of 0º C and a stopping distance of 80 feet, the correlation would
   A. Decrease, since this new point is an outlier that does not follow the pattern in the data.
   B. Increase, since this new point is an outlier that does not follow the pattern in the data.
   C. Stay nearly the same, since correlation is resistant to outliers.
   D. Increase, since there would be more data points.
   E. Whether this data point causes and increase or decrease cannot be determined without recalculating the correlation.

84. Use Scenario 3-3. If the stopping distance were measured in meters rather than feet (1 meter = approx. 3.28 feet), how would the correlation r change?
   A. r would be smaller, since the same distances are smaller when measured in meters.
   B. r would be larger, since the same distances are smaller when measured in meters.
   C. r would not change, since the calculation of r does not depend on the units used.
   D. r would not change, because only changes in the units of the x-variable (temperature, in this case. can influence the value of r.
   E. r could be larger or smaller—we can’t tell without recalculating correlation.

85. Which of the following is true of the correlation r?
   A. It is a resistant measure of association.
   B. \(-1 = r = 1\).
   C. If r is the correlation between X and Y, then -r is the correlation between Y and X.
   D. Whenever all the data lie on a perfectly straight-line, the correlation r will always be equal to +1.0.
   E. All of the above.
86. Consider the following scatter plot of two variables, X and Y.

We may conclude that the correlation between X and Y
A. must be close to −1, since the relationship is between X and Y is clearly non-linear.
B. must be close to 0, since the relationship is between X and Y is clearly non-linear.
C. is close to 1, even though the relationship is not linear.
D. may be exactly 1, since all the points line of the same curve.
E. greater than 1, since the relationship is non-linear.

87. Which one of the following statements is correct?
A. Faculty who are good researchers tend to be poor teachers and vice versa, so the correlation between teaching and research is 0.
B. Women tend to be, on average, about 3.5 inches shorter than the men they marry, so the correlation between the heights of spouses must be negative.
C. A researcher finds the correlation between the shoe size of children and their score on a reading test to be 0.22. The researcher must have made a mistake since these two variables are clearly unrelated and must have correlation 0.
D. If people with larger heads tend to be more intelligent, then we would expect the correlation between head size and intelligence to be positive.
E. The correlation \( r \) equals the proportion of times that two variables lie on a straight-line.

88. Which of the following best describes the correlation \( r \)?
A. The average of the products of each of the X and Y values for each point
B. The average of the products of the standardized scores of X and Y for each point.
C. The average of the squared products of the standardized scores of X and Y for each point.
D. The average of the differences between each X value and each Y value.
E. The average perpendicular distance between each data point and the least-squares regression line.

89. Consider the scatter plot below for a very small data set, consisting of the heights of five fathers (x) and their sons (y). The “M” in the plot indicates the point \((\bar{x}, \bar{y})\). The letters A – E are labels for the five father-son pairs.
Which father-son pair contributes the largest positive quantity to the correlation between father and son heights?
A. Pair A  
B. Pair B  
C. Pair C  
D. Pair D  
E. Pair E

90. The scatter plot below describes the relationship between heights of 36 students and the number of words they spelled correctly in a spelling bee. The closed circles represent first graders and the open circles represent fifth graders.

Which of the following statements is not supported by the information in the scatter plot?
A. Most of the fifth graders spelled more words correctly than most of the first graders.
B. When the data for first and fifth grades is combined, there is a moderately strong positive relationship between height and how many words were spelled correctly.
C. When the two grades are considered separately, there is little or no relationship between height and how many words were spelled correctly.
D. The tallest first grader spelled more words correctly than five of the fifth graders.
E. All of the fifth graders are taller than the tallest first grader.

Scenario 3-4
Consider the following scatterplot of amounts of CO (carbon monoxide) and NOX (nitrogen oxide) in grams per mile driven in the exhausts of cars. The least-squares regression line has been drawn in the plot.

___ 91. Use Scenario 3-4. The intercept of the least-squares regression line is approximately
A. −0.7.
B. −0.1.
C. 1.8.
D. 2.0.
E. 18.

___ 92. Use Scenario 3-4. Based on the scatterplot, the least-squares line would predict that a car that emits 10 grams of CO per mile driven would emit approximately how many grams of NOX per mile driven?
A. 10.0
B. 1.7
C. 2.2
D. 1.1
E. 0.7

___ 93. Use Scenario 3-4. In the scatterplot, the point indicated by the open circle
A. has a negative value for the residual.
B. has a positive value for the residual.
C. has a zero value for the residual.
D. has a zero value for the correlation.
E. is an outlier.

---

**Scenario 3-5**
In a statistics course a linear regression equation was computed to predict the final exam score from the score on the first test. The equation of the least-squares regression line was \( \hat{y} = 10 + 0.9x \) where \( \hat{y} \) represents the predicted final exam score and \( x \) is the score on the first exam.

___ 94. Use Scenario 3-5. The first test score is
A. the intercept.
B. the slope.
C. the explanatory variable.
D. the response variable.
95. Use Scenario 3-5. Suppose Joe scores a 90 on the first exam. What would be the predicted value of his score on the final exam?
A. 91
B. 90
C. 89
D. 81
E. Cannot be determined from the information given. We also need to know the correlation.

96. “Least-squares” in the term “least-squares regression line” refers to
A. Minimizing the sum of the squares of all values of the explanatory variable.
B. Minimizing the sum of the squares of all values of the response variable.
C. Minimizing the products of each value of the response variable and the predicted value based on the regression equation.
D. Minimizing the sum of the squares of the residuals.
E. Minimizing the squares of the differences between each value of the response variable and each value of the explanatory variable.

97. Which of the following statements are true about the least-squares regression line?
I. The slope is the predicted change in the response variable associated with a unit increase in the explanatory variable.
II. The line always passes through the point \((J, M)\), the means of the explanatory and response variables, respectively.
III. It is the line that minimizes the sum of the squared residuals.
A. I only.
B. II only.
C. III only.
D. I and III only.
E. I, II, and III are all true.

Scenario 3-6
A researcher wishes to study how the average weight \(Y\) (in kilograms) of children changes during the first year of life. He plots these averages versus the age \(X\) (in months) and decides to fit a least-squares regression line to the data with \(X\) as the explanatory variable and \(Y\) as the response variable. He computes the following quantities.
\[ r = \text{correlation between } X \text{ and } Y = 0.9 \]
\[ J = \text{mean of the values of } X = 6.5 \]
\[ M = \text{mean of the values of } Y = 6.6 \]
\[ Sx = \text{standard deviation of the values of } X = 3.6 \]
\[ Sy = \text{standard deviation of the values of } Y = 1.2 \]

98. Use Scenario 3-6. The slope of the least-squares line is
A. 0.30.
B. 0.88.
C. 1.01.
D. 3.0.
E. 2.7.

99. Use Scenario 3-6. The y-intercept of the least-squares line is
A. \(-10.95\)
B. 4.52
C. 4.65
D. 8.48
E. 8.55

100. The fraction of the variation in the values of a response $y$ that is explained by the least-squares regression of $y$ on $x$ is the
A. correlation coefficient.
B. slope of the least-squares regression line.
C. square of the correlation coefficient.
D. intercept of the least-squares regression line.
E. sum of the squared residuals.

101. The correlation between the age and height of children is found to be about $r = 0.7$. Suppose we use the age $x$ of a child to predict the height $y$ of the child. We conclude that
A. the least-squares regression line of $y$ on $x$ would have a slope of 0.7.
B. the fraction of the variation in heights explained by the least-squares regression line of $y$ on $x$ is 0.49.
C. about 70% of the time, age will accurately predict height.
D. the fraction of the variation in heights explained by the least-squares regression line of $y$ on $x$ is 0.70.
E. the line explains about 49% of the data.

102. Which of the following is correct?
A. The correlation $r$ is the slope of the least-squares regression line.
B. The square of the correlation is the slope of the least-squares regression line.
C. The square of the correlation is the proportion of the data lying on the least-squares regression line.
D. The mean of the residuals from least-squares regression is 0.
E. The sum of the squared residuals from the least-squares line is 0.

103. Suppose we fit the least-squares regression line to a set of data. If a plot of the residuals shows a curved pattern,
A. a straight line is not a good summary for the data.
B. the correlation must be 0.
C. the correlation must be positive.
D. outliers must be present.
E. $r^2 = 0$.

104. If removing an observation from a data set would have a marked change on the equation of the least-squares regression line, the point is called
A. resistant.
B. a residual.
C. influential.
D. a response.
E. an outlier.

105. Which of the following statements about influential points and outliers are true?
I. An influential point always has a high residual.
II. Outliers are always influential points.
III. Removing an influential point always causes a marked change in either the correlation, the regression equation, or both.
A. I only.
B. II only.
C. III only.
D. II and III only.
E. I, II, and III are all true.

106. Suppose a straight line is fit to data having response variable $y$ and explanatory variable $x$. Predicting values of $y$ for values of $x$ outside the range of the observed data is called
A. contingency.
B. extrapolation.
C. causation.
D. correlation.
E. interpolation.

Scenario 3-7

Below is a scatter plot (with the least squares regression line) for calories and protein (in grams) in one cup of 11 varieties of dried beans. The computer output for this regression is below the plot.

107. Use Scenario 3-7. Which of the following statements is a correct interpretation of the slope of the regression line?
A. For each 1-unit increase in the calorie content, the predicted protein content increases by 2.08 grams.
B. For each 1-unit increase in the calorie content, the predicted protein content increases by 0.063 grams.
C. For each 1-gram increase in the protein content, the predicted calorie content increases by 2.08 grams.
D. For each 1-gram increase in the protein content, the predicted calorie content increases by 0.063 grams.
E. For each 1-gram increase in the protein content, the predicted calorie content increases by 0.024 grams.

108. Use Scenario 3-7. Which of the following best describes what the number $S = 3.37648$ represents?
A. The slope of the regression line is 3.37648.
B. The standard deviation of the explanatory variable, calories, is 3.37648.
C. The standard deviation of the response variable, protein content, is 3.37648.
D. The standard deviation of the residuals is 3.37648.
E. The ratio of the standard deviation of protein to the standard deviation of calories is 3.37648.

109. Use Scenario 3-7. The circled point on the scatter plot represents lima beans, which have 621 calories and 37 grams of protein. The residual for lima beans is:
A. –37.0
B. –4.18
C. 4.18
D. 37.0
E. 41.18

110. Use Scenario 3-7. One cup of dried soybeans contains 846 calories. Which of the following statements is appropriate?
A. We can predict that the protein content for soybeans is 55.4 grams.
B. We can predict that the protein content for soybeans is 53.3 grams
C. We can predict that the protein content for soybeans is 51.2 grams
D. Unless we are given the observed protein content for soybeans, we can’t calculate the predicted protein content.
E. It would be inappropriate to predict the protein content of soybeans with this regression model, since their calorie content is well beyond the range of these data.

111. The least-squares regression line is fit to a set of data. If one of the data points has a positive residual, then
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 is probably an influential point.
E. all of the above.

112. Which of the following statements concerning residuals is true?
A. The sum of the residuals is always 0.
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.
D. An influential point on a scatterplot is not necessarily the point with the largest residual.
E. All of the above.

113. A study of child development measures the age (in months) at which a child begins to talk and also the child's score on an ability test given several years later. The study asks whether the age at which a child talks helps predict the later test score. The least-squares regression line of test score \( y \) on age \( x \) is \( y = 110 - 1.3x \). According to this regression line, what happens (on the average) to children who talk one month later than other children?
A. Their predicted test scores go down 110 points.
B. Their predicted test scores go down 1.3 points.
C. Their predicted test scores go up 110 points.
D. Their predicted test scores go up 1.3 points.
E. Their predicted test scores are 108.7.

Scenario 3-8
A fisheries biologist studying whitefish in a Canadian Lake collected data on the length (in centimeters) and egg production for 25 female fish. A scatter plot of her results and computer regression analysis of egg production versus fish length are given below. *Note that Number of eggs is given in thousands (i.e., “40” means 40,000 eggs).*

**Egg production vs fish length**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-142.74</td>
<td>25.55</td>
<td>-5.59 0.000</td>
<td></td>
</tr>
<tr>
<td>Fish length</td>
<td>39.25</td>
<td>5.392</td>
<td>7.28  0.000</td>
<td></td>
</tr>
</tbody>
</table>

\[ S = 6.75133 \quad \text{R-Sq} = 69.7\% \quad \text{R-Sq(adj)} = 68.4\% \]

114. Use Scenario 3-8. The equation of the least-squares regression line is
   A. Eggs = -142.74 + 39.25(Length)
   B. Eggs = 39.25 - 142.74(Length)
   C. Eggs = 25.55 + 5.392(Length)
   D. Eggs = 25.55 + 5.392(Eggs)
   E. Eggs = -142.74 + 39.25(Eggs)

115. Use Scenario 3-8. On average, how far are the predicted y-values from the actual y-values?
   A. 25.55
   B. 5.392
   C. 6.75133
   D. 0.697
   E. Cannot be determined without the original data.

116. Use Scenario 3-8. Which of the following statements can be made on the basis of the computer output?
   A. 83.5% of the variation in egg production can be accounted for by the linear regression of egg production on fish length.
   B. 69.7% of the variation in egg production can be accounted for by the linear regression of egg production on fish length.
   C. 83.5% of the variation in fish length can be accounted for by the linear regression of egg production on fish length.
   D. 69.7% of the variation in fish length can be accounted for by the linear regression of egg production on fish length.
   E. 68.4% of the variation in fish length can be accounted for by the linear regression of egg production on fish length.
117. Use Scenario 3-8. Which of the following is the plot of residuals versus fish lengths?

A. ![Residuals versus fish length](image1)

B. ![Residuals versus fish length](image2)

C. ![Residuals versus fish length](image3)

D. ![Residuals versus fish length](image4)

E. ![Residuals versus fish length](image5)

118. A study of the effects of television measured how many hours of television each of 125 grade school children watched per week during a school year and their reading scores. The study found that children who watch more television tend to have lower reading scores than children who watch fewer hours of television. The study report says that "Hours of television watched explained 9% of the observed variation in the reading scores of the 125 subjects." The correlation between hours of TV and reading score must be

A. \( r = 0.09 \).

B. \( r = -0.09 \).

C. \( r = 0.3 \).

D. \( r = -0.3 \).

E. Can't tell from the information given.

**Scenario 3-9**
A study gathers data on the outside temperature during the winter, in degrees Fahrenheit, and the amount of natural gas a household consumes, in cubic feet per day. Call the temperature $x$ and gas consumption $y$. The house is heated with gas, so $x$ helps explain $y$. The least-squares regression line for predicting $y$ from $x$ is:

$$\hat{y} = 1344 - 19x$$

119. Use Scenario 3-9. On a day when the temperature is 20°F, the regression line predicts that gas used will be about
A. 1724 cubic feet.
B. 1383 cubic feet.
C. 1325 cubic feet.
D. 964 cubic feet.
E. none of the above.

120. Use Scenario 3-9. When the temperature goes up 1 degree, what happens to the gas usage predicted by the regression line?
A. It goes up 1 cubic foot.
B. It goes down 1 cubic foot.
C. It goes up 19 cubic feet.
D. It goes down 19 cubic feet.
E. Can't tell without seeing the data.

121. Use Scenario 3-9. What does the number 1344 represent in the equation?
A. Predicted gas usage (in cubic feet) when the temperature is 19 degrees Fahrenheit.
B. Predicted gas usage (in cubic feet) when the temperature is 0 degrees Fahrenheit.
C. It's the $y$-intercept of the regression line, but it has no practical purpose in the context of the problem.
D. The maximum possible gas a household can use.
E. None of the above.

122. Students with above-average scores on Exam 1 in STAT 001 tend to also get above-average scores on Exam 2. But the relationship is only moderately strong. In fact, a linear relationship between Exam 2 scores and Exam 1 scores explains only 36% of the variance of the Exam 2 scores.
A. The correlation between Exam 1 scores and Exam 2 scores is $r = .36$.
B. The correlation between Exam 1 scores and Exam 2 scores is $r = .6$.
C. The correlation between Exam 1 scores and Exam 2 scores is $r = \pm .36$ (can't tell which).
D. The correlation between Exam 1 scores and Exam 2 scores is $r = \pm .6$ (can't tell which).
E. There is not enough information to say what $r$ is.

123. Scores on the 1995 SAT verbal aptitude test $x$ among Kentucky high school seniors were normally distributed with mean 420 and standard deviation 80. Scores on the 1995 SAT quantitative aptitude test $y$ among Kentucky high school seniors were normally distributed with mean 440 and standard deviation 60. The least-squares regression line has the equation $y = .6x + 188$. The correlation between verbal scores and math scores is
A. –.8
B. 0
C. .45
D. .8
E. cannot be determined from the information given
124. You are examining the relationship between \( x = \) the height of red oak trees and \( y = \) the number of acorns produced in a five year period. You calculate a correlation coefficient and a least-squares regression line of \( y \) on \( x \). If you switched the variables (that is, let \( x = \) number of acorns and \( y = \) height of trees), which of the following would be true?
A. both the correlation coefficient and the regression line would be unchanged.
B. the correlation coefficient would change, but the regression line would not change.
C. the correlation coefficient would not change, but the regression line would change.
D. neither the correlation coefficient nor the regression equation would change.
E. only the \( y \)-intercept of the regression line would change, the slope of the line and the correlation coefficient would not change.

125. Which of the following statements describes what the standard deviation of residuals for a regression equation can be used for?
I. It describes the typical vertical distance between an observed data point and the regression line.
II. It evaluates whether a linear model is appropriate for a set of data.
III. It measures the overall precision of predictions made using the regression equation.
A. I only
B. II only
C. III only
D. Both I and II
E. Both I and III

**Scenario 4-1**
A sportswriter wants to know how strongly Lafayette residents support the local minor league baseball team, the Lafayette Leopards. She stands outside the stadium before a game and interviews the first 20 people who enter the stadium.

126. Use Scenario 4-1. The intended *population* for this survey is
A. all residents of Lafayette.
B. all Leopard fans.
C. all people attending the game the day the survey was conducted.
D. the 20 people who gave the sportswriter their opinion.
E. all American adults.

127. Use Scenario 4-1. The *sample* for the survey is
A. all residents of Lafayette.
B. all Leopard fans.
C. all people attending the game the day the survey was conducted.
D. the 20 people who gave the sportswriter their opinion.
E. the sportswriter.

128. Use Scenario 4-1. The newspaper asks you to comment on their survey of local opinion. You say:
A. This is a simple random sample. It gives very accurate results.
B. This is a simple random sample. The results are not biased, but the sample is too small to have high precision.
C. This is a census, because all fans had a chance to be asked. It gives very accurate results.
D. This is a convenience sample. It will almost certainly overestimate the level of support among all Lafayette residents.
E. This is a convenience sample. It will almost certainly underestimate the level of support among all Lafayette residents.
Scenario 4-2
You want to know the opinions of American school teachers about establishing a national test for high school graduation. You obtain a list of the members of the National Education Association (the largest teachers' union) and mail a questionnaire to 2500 teachers chosen at random from this list. In all 1347 teachers return the questionnaire.

129. Use Scenario 4-2. The sampling frame is
   A. the 1347 teachers who mail back the questionnaire.
   B. the 2500 teachers to whom you mailed the questionnaire.
   C. all members of the National Education Association.
   D. all American school teachers.
   E. all American school students.

130. Use Scenario 4-2. The sample is
   A. the 1347 teachers who mail back the questionnaire.
   B. the 2500 teachers to whom you mailed the questionnaire.
   C. all members of the National Education Association.
   D. all American school teachers.
   E. all American school students.

131. A study sponsored by American Express Co. and the French government tourist office found that old stereotypes about French unfriendliness were not true. The respondents were more than 1000 Americans who have visited France more than once for pleasure over the past two years. The results of this study are probably
   A. very accurate, given the large sample size.
   B. very inaccurate because the sample is only a small fraction of all Americans who have visited France.
   C. extremely variable, because people's opinions differ so greatly.
   D. biased, overstating the extent to which the old stereotypes were not true.
   E. biased, understating the extent to which the old stereotypes were not true.

132. A candidate for mayor of Dallas calls 1,000 people chosen at random from the city telephone directory; 850 of them respond. What are the sampling frame and the sample in this example?
   A. Sampling frame: the telephone directory. Sample: the 850 people who respond.
   B. Sampling frame: the telephone directory. Sample: the 1,000 people who are called.
   C. Sampling frame: the 1,000 people who are called. Sample: the 850 people who respond.
   D. Sampling frame: all Dallas residents. Sample: the 1,000 people who are called.
   E. Sampling frame: all Dallas residents. Sample: the 850 people who respond.

133. In order to assess the opinion of students at the University of Minnesota on campus snow removal, a reporter for the student newspaper interviews the first 12 students he meets who are willing to express their opinion. The method of sampling used is
   A. a census
   B. a systematic sample
   C. a voluntary sample
   D. a convenience sample
   E. a simple random sample

134. A television station is interested in predicting whether voters in its viewing area are in favor of offshore drilling. It asks its viewers to phone in and indicate whether they support/are in favor of or are opposed to this practice. Of the 2241 viewers who phoned in, 1574 (70%) were opposed to offshore drilling. The viewers who phoned in are
   A. a voluntary response sample.
B. a convenience sample.
C. a probability sample.
D. a population.
E. a simple random sample.

135. A poll conducted by the student newspaper asked, "Who do you believe will win the Ohio State Undergraduate Student Government elections?" In order to vote, one had to access the student newspaper's Web site and record one's vote. The results of the poll were summarized in a graphic similar to the following.

Which of the following statements is true about these results?
A. The results of the survey are unreliable because response to the survey was voluntary.
B. The sample is large enough to eliminate potential sources of bias in the design of the poll.
C. This is not an appropriate way of presenting the results—a bar graph should have been used instead.
D. Patel and Patel have such a large majority that, even though there are flaws in the poll, they are still almost certain to win.
E. There must be an error. These percentages aren't possible.

136. A news release for a diet products company reports: "There's good news for the 65 million Americans currently on a diet." Its study showed that people who lose weight can keep it off. The sample was twenty graduates of the company's program who endorse it in commercials. The results of the study are probably
A. biased, overstating the effectiveness of the diet.
B. biased, understating the effectiveness of the diet.
C. unbiased because these are nationally recognized individuals.
D. unbiased, but they could be more accurate. A larger sample size should be used.
E. biased, but it is hard to tell whether the results will overstate or understate the effects of the diet.

137. A simple random sample of size $n$ is defined to be
A. a sample of size $n$ chosen in such a way that every unit in the population has the same chance of being selected.
B. a sample of size $n$ chosen in such a way that every unit in the population has a known nonzero chance of being selected.
C. a sample of size $n$ chosen in such a way that every set of $n$ units in the population has an equal chance to be the sample actually selected.
D. a sample of size $n$ chosen in such a way that each selection is made independent of every other selection.
E. all of the above. They are essentially identical definitions.
138. A marketing research firm wishes to determine if the adult men in Laramie, Wyoming, would be interested in a new upscale men’s clothing store. From a list of all residential addresses in Laramie, the firm selects a simple random sample of 100 and mails a brief questionnaire to each. The chance that all 100 homes in a particular neighborhood in Laramie end up being the sample of residential addresses selected is
A. the same as for any other set of 100 residential addresses.
B. exactly 0. Simple random samples will spread out the addresses selected.
C. reasonably large due to the “cluster” effect.
D. 100 divided by the size of the population of Laramie.
E. large since the population of Laramie is small.

139. A simple random sample is
A. any sample selected by using chance.
B. any sample that gives every individual the same chance to be selected.
C. a sample that gives every possible sample of the same size the same chance to be selected.
D. a sample that selects equal numbers of individuals from each stratum.
E. a sample that contains the same percent of each subgroup in the population.

140. Simple random sampling
A. reduces bias resulting from poorly worded questions.
B. offsets bias resulting from undercoverage and nonresponse.
C. reduces bias resulting from the behavior of the interviewer.
D. reduces variability.
E. None of the above.

141. An example of a nonsampling error that can reduce the accuracy of a sample survey is
A. using voluntary response to choose the sample.
B. using the telephone directory as the sampling frame.
C. interviewing people at shopping malls to obtain a sample.
D. variation due to chance in choosing a sample at random.
E. many members of the sample cannot be contacted.

Scenario 4-3
We wish to choose a simple random sample of size three from the following employees of a small company. To do this, we will use the numerical labels attached to the names below.


We will also use the following list of random digits, reading the list from left to right, starting at the beginning of the list.

11793  20495  05907  11384  44982  20751  27498  12009  45287  71753  98236  66419  84533

142. Use Scenario 4-3. The simple random sample is
A. 117.
B. Bechhofer, Bechhofer again, and Taylor.
C. Bechhofer, Taylor, Weiss.
D. Kesten, Kiefer, Taylor.
E. Taylor, Weiss, Ito.
143. Use Scenario 4-3. Which of the following statements is true?
   A. If we use another list of random digits to select the sample, we would get the same result as that obtained with the list actually used.
   B. If we use another list of random digits to select the sample, we would get a completely different sample than that obtained with the list actually used.
   C. If we use another list of random digits to select the sample, we would get, at most, one name in common with that obtained with the list actually used.
   D. If we use another list of random digits to select the sample, the result obtained with the list actually used would be just as likely to be selected as any other set of three names.
   E. If we use another list of random digits to select the sample, the result obtained with the list actually used would be far less likely to be selected than any other set of three names.

144. Use Scenario 4-3. Which of these statements about the table of random digits is true?
   A. Every row must have exactly the same number of 0's and 1's.
   B. In the entire table, there are exactly the same number of 0's and 1's.
   C. If you look at 100 consecutive pairs of digits anywhere in the table, exactly 1 pair is 00.
   D. All of these are true.
   E. None of these is true.

145. The eight students listed below are enrolled in a new honors course developed by the chemistry department.
   1. Alvarez
   2. Barlow
   3. Nahhas
   4. Salter
   5. Miller
   6. Pfouts
   7. Berliner
   8. Verducci

Starting at the beginning of the random number list below, choose a simple random sample of four students to be interviewed in detail about the quality of the course. Use the labels attached to the eight names.

41842 81868 71035 09001 43367 49497 54580 81507

The sample you obtain is
   A. 4, 1, 8, and 4.
   B. Alvarez, Barlow, Nahhas, and Salter.
   C. Alvarez, Barlow, Salter, and Verducci.
   D. Salter, Alvarez, Verduci, Salter.
   E. Salter, Alvarez, Verduci, Pfouts.

**Scenario 4-4**
You want to take an SRS of 50 of the 816 students who live in a dormitory on campus. You label the students 001 to 816 in alphabetical order. In the table of random digits you read the entries.

95592 94007 69769 33547 72450 16632 81194 14873

146. Use Scenario 4-4. The first three students in your sample have labels
   A. 955, 929, 400.
   B. 400, 769, 769.
   C. 559, 294, 007.
   D. 929, 400, 769.
   E. 400, 769, 335.

147. Use Scenario 4-4. Another correct choice of labels for the 816 students is
148. A public opinion poll in Ohio wants to determine whether or not registered voters in the state approve of a measure to ban smoking in all public areas. They select a simple random sample of fifty registered voters from each county in the state and ask whether they approve or disapprove of the measure. This is an example of a
A. systematic random sample.
B. stratified random sample.
C. multistage sample.
D. simple random sample.
E. cluster sample.

149. A stratified random sample is appropriate when
A. It is impractical to take a simple random sample because the population is too large.
B. The population can be easily subdivided into groups according to some categorical variable, and the variable you are measuring is quite different within the groups but very similar between groups.
C. The population can be easily subdivided into groups according to some categorical variable, and the variable you are measuring is very similar within the groups but quite different between groups.
D. You intend to take a sample of more than 100 individuals.
E. You want to avoid undercoverage of certain groups.

150. A stratified random sample addresses the same issues as which of the following experimental designs?
A. A block design.
B. A double-blind experiment.
C. An experiment with a placebo.
D. A matched pairs design.
E. A confounded, nonrandomized study.

151. We divide the class into two groups: first year students and others. We then take random samples from each group. This is an example of
A. simple random sampling
B. cluster sampling
C. multistage sampling
D. stratified random sampling
E. systematic random sampling

152. To determine the proportion of each color of Peanut Butter M&M, you buy 10 1.69 ounce packages and count how many there are of each color. This is an example of
A. simple random sampling
B. cluster sampling
C. multistage sampling
D. stratified random sampling
E. systematic random sampling
153. In order to select a sample of undergraduate students in the United States, I select a simple random sample of four states. From each of these states, I select a simple random sample of two colleges or universities. Finally, from each of these eight colleges or universities, I select a simple random sample of 20 undergraduates. My final sample consists of 160 undergraduates. This is an example of
A. simple random sampling.
B. stratified random sampling.
C. multistage sampling.
D. convenience sampling.
E. cluster sampling.

154. An opinion research firm wants to find the country’s reaction to a speech by a famous politician. They randomly select six states, then randomly select ten Zip Codes from each state. Fifty people from each Zip Code are randomly selected for the survey. This is an example of
A. convenience sampling.
B. cluster sampling.
C. stratified random sampling.
D. simple random sampling.
E. multistage sampling.

155. You plan to give a math achievement test to samples of 15 year-olds from both the U.S. and Korea in order to compare mathematics knowledge in the two countries. In each country, you will randomly choose

300 students from low-income families
400 students from middle-income families
200 students from high-income families

The sample from Korea is a
A. multistage sample.
B. simple random sample.
C. convenience sample.
D. voluntary response sample.
E. stratified random sample.

156. A marine biologist wants to estimate the mean size of the barnacle *Semibalanus balnoides* on a stretch of rocky shoreline. To do so, he randomly selected twenty 10-cm. square plots and measured the size of every barnacle in each plot. This is an example of
A. convenience sampling.
B. cluster sampling.
C. stratified random sampling.
D. simple random sampling.
E. multistage sampling.

157. A 1992 Roper poll found that 22% of Americans say that the Holocaust may not have happened. The actual question asked in the poll was “*Does it seem possible or impossible to you that the Nazi extermination of the Jews never happened?*” and 22% responded possible. The results of this poll cannot be trusted because
A. undercoverage is present. Obviously, those people who did not survive the Holocaust could not be in the poll.
B. the question is worded in a confusing manner.
C. we do not know who conducted the poll or who paid for the results.
D. nonresponse is present. Many people will refuse to participate, and those who do will be biased in their opinions.
E. the question is clearly biased in the direction of a "possible" answer.
158. In the late 1990’s Scotland was considering independence from England. An opinion poll showed that 51% of Scots favor "independence." Another poll taken at the same time showed that only 34% favored being "separate" from England. The reason these results differ by so much is that
A. samples will usually differ just by chance due to random sampling.
B. the wording of questions has a big effect on poll results.
C. more follow-up efforts reduced the nonresponse rate of the second poll.
D. the sample sizes are different, so the margins of error are different.
E. the second poll suffered from undercoverage.

159. A local tax reform group polls the residents of the school district and asks the question, “Do you think the school board should stop spending taxpayers’ money on non-essential arts programs in elementary schools?” The results of this poll are likely to
A. Underestimate support for arts programs because of undercoverage.
B. Underestimate support for arts programs because of nonsampling error.
C. Overestimate support for arts programs because of undercoverage.
D. Overestimate support for arts programs because of nonsampling error.
E. Accurately estimate support for arts programs.

160. Frequently, telephone poll-takers call near dinner time—between 6 pm and 7 pm—because most people are at home then. This is an effort to avoid
A. voluntary response bias.
B. calling people after they have gone to bed.
C. a convenience sample.
D. nonresponse.
E. response bias.

161. The Bradley effect is a theory proposed to explain observed discrepancies between voter opinion polls and election outcomes in some elections where a white candidate and a non-white candidate run against each other. The theory proposes that some voters tend to tell pollsters that they are undecided or likely to vote for a non-white candidate, and yet, on election day, vote for the white opponent. This is an example of
A. voluntary response bias.
B. bias resulting from question wording.
C. undercoverage.
D. nonresponse.
E. response bias.

162. Just before the presidential election of 1936, the magazine Literary Digest predicted—incorrectly, as it turned out—that Alf Landon would defeat Franklin Delano Roosevelt. Landon lost in a landslide. It turned out that the magazine had only polled its own subscribers, plus others from a list of automobile owners and a list of people who had telephone service. All three groups had higher than typical incomes during the Great Depression. This is an example of
A. voluntary response bias.
B. bias resulting from question wording.
C. undercoverage.
D. nonresponse.
E. response bias.

163. Which of the following is not a major principle of good design for all experiments?
A. Comparison to a control.
B. Replication
C. Blocking
D. Randomization
E. All of these are important principles for every experiment.

___ 164. The essential difference between an experiment and an observational study is that
A. observational studies may have confounded variables, but experiments never do.
B. in an experiment, people must give their informed consent before being allowed to participate.
C. observational studies are always biased.
D. observational studies cannot have response variables.
E. an experiment imposes treatments on the subjects, but an observational study does not.

Scenario 4-5
In order to assess the effects of exercise on reducing cholesterol, a researcher took a random sample of fifty people from a local gym who exercised regularly and another random sample of fifty people from the surrounding community who did not exercise regularly. They all reported to a clinic to have their cholesterol measured. The subjects were unaware of the purpose of the study, and the technician measuring the cholesterol was not aware of whether or not subjects exercised regularly.

___ 165. Use Scenario 4-5. This is a(n)
A. observational study.
B. experiment, but not a double blind experiment.
C. double blind experiment.
D. matched pairs experiment.
E. block design.

___ 166. Use Scenario 4-5. Which of the following best describes the inferences the researcher can make based in his results?
A. He can make inferences about cause and effect, but not about the populations from which the samples were taken.
B. He can make inferences about the populations from which the samples were taken, but not about cause and effect.
C. He can make inferences about both cause and effect and the populations from which the samples were taken.
D. He cannot make inferences about either cause and effect or the populations from which the samples were taken.
E. There is not enough information to make judgments about the scope of inference.

___ 167. A market research company wishes to find out whether the population of students at a university prefers brand A or brand B of instant coffee. A random sample of students is selected, and each one is asked to try brand A first and then brand B (or vice versa, with the order determined at random). They then indicate which brand they prefer. The response variable is
A. whether brand A or B is tried first.
B. which brand they prefer.
C. the two brands of coffee.
D. the identity of the student.
E. none of these.

___ 168. The most important advantage of experiments over observational studies is that
A. experiments are usually easier to carry out.
B. experiments can give better evidence of causation.
C. confounding cannot happen in experiments.
D. an observational study cannot have a response variable.
E. observational studies cannot use random samples.

**Scenario 4-6**

Does caffeine improve exam performance? Suppose all students in the 8:30 section of a course are given a "treatment" (two cups of coffee) and all students in the 9:30 section are not permitted to have any caffeine before a mid-term exam.

169. Use Scenario 4-6. Unfortunately, any systematic difference between the two sections on the exam might be due to the fact that the 8:30 and 9:30 classes have different instructors. This is an example of
A. placebo effect.
B. bias.
C. confounding.
D. observational study.
E. stratification.

170. Use Scenario 4-6. The response variable in this study is
A. two cups of coffee.
B. the time the class is held.
C. class attendance.
D. teacher's performance.
E. exam performance.

171. Use Scenario 4-6. Instead of giving all students in the 8:30 section two cups of coffee, students in the 8:30 section are randomly assigned to a treatment group (two cups of coffee) or a control group (two cups of decaffeinated coffee). The coffee is so bad that students cannot tell whether they are in the treatment or the control group. As it turns out, students in both groups do better on the exam than students in the 9:30 section, who weren't given anything. This could be the result of
A. the placebo effect.
B. an observational study.
C. voluntary response.
D. sampling variability.
E. all of the above.

172. Use Scenario 4-6. Suppose half of the 8:30 students are randomly allocated to the treatment group (two cups of coffee), the other half to the control group (two cups of decaf). In addition, half of the 9:30 students are randomly allocated to the treatment group, the other half to the control group. This is an example of a
A. voluntary response study.
B. stratified sampling procedure.
C. matched pairs design.
D. completely randomized design.
E. block design.

**Scenario 4-7**

A farmer wishes to determine which of two brands of baby pig pellets, Kent or Moormans, produces better weight gains. Two of his sows each give birth to litters of 10 pigs on the same day, so he decides to give the baby pigs in litter A only Kent pellets, while the pigs in litter B will get only Moormans pellets. After four weeks, the average weight gain for pigs in litter A is greater than the average weight gain for pigs in litter B.

173. Use Scenario 4-7. The farmer has conducted a(n)
A. stratified random sample.
B. matched pairs design.
C. observational study.
D. experiment, but not a completely randomized experiment.
E. completely randomized experiment.

174. Use Scenario 4-7. The brand of pellets is
   A. a parameter.
   B. the response variable.
   C. the explanatory variable.
   D. the placebo effect.
   E. a lurking variable.

175. Use Scenario 4-7. The feed they get is not the only factor affecting the rate at which pigs gain weight. Genetic differences also affect weight gain. It is likely that the pigs in litter A are genetically different from the pigs in litter B, since the two litters have different mothers. Since the farmer is only interested in determining which brand of pellets is better, the study suffers from
   A. confounding.
   B. common response.
   C. convenience sampling.
   D. invalid measurement.
   E. experimenter bias.

176. Use Scenario 4-7. If the farmer had fed Kent pellets to an SRS of 5 pigs from litter A and an SRS of 5 pigs from litter B, with the remaining 10 pigs getting Moormans pellets, then he would have been using
   A. a systematic random sample.
   B. a convenience sample.
   C. a matched-pairs design.
   D. a block design.
   E. a completely randomized design.

177. Use Scenario 4-7. The weight of the pigs after four weeks is
   A. a parameter.
   B. the response variable.
   C. the explanatory variable.
   D. the placebo effect.
   E. stratified.

178. An experiment was conducted by some students to explore the nature of the relationship between a person’s heart rate (measured in beats per minute) and the frequency at which that person stepped up and down on steps of various heights. Three rates of stepping and two different step heights were used. A subject performed the activity (stepping at one of the three stepping rates at one of the two possible heights) for three minutes. Heart rate was then measured at the end of this period. The variables “stepping rate” and “step height” are the
   A. factors.
   B. levels.
   C. controls.
   D. units.
   E. response variables.

**Scenario 4-8**
Researchers wish to determine if a new experimental medication will reduce the symptoms of allergy sufferers without the side effect of drowsiness. To investigate this question, the researchers randomly assigned 100 adult volunteers who suffer from allergies to two groups. They gave the new medication to the subjects in one group and an existing medication to the subjects in the other group. Forty-four percent of those in the treatment group and 28% of those in the control group reported a significant reduction in their allergy symptoms without any drowsiness.

___ 179. Use Scenario 4-8. The experimental units are the
A. researchers.
B. 100 adult volunteers.
C. all the volunteers who reported a significant reduction in their allergy symptoms without any drowsiness.
D. all the volunteers who did not report a significant reduction in their allergy symptoms without any drowsiness.
E. pills containing the new experimental medication.

___ 180. Use Scenario 4-8. Which of the following best describes the inferences the researchers can make based in his results?
A. They can make inferences about cause and effect, but not about the populations from which the samples were taken.
B. They can make inferences about the populations from which the samples were taken, but not about cause and effect.
C. They can make inferences about both cause and effect and the populations from which the samples were taken.
D. They cannot make inferences about either cause and effect or the populations from which the samples were taken.
E. There is not enough information to make judgments about the scope of inference.

___ 181. Medical researchers are excited about a new cancer treatment that destroys tumors by cutting off their blood supply. To date, the treatment has only been tried on mice, but in mice it has been nearly 100% effective in eradicating tumors and appears to have no side effects. As evidence of the effectiveness of the new treatment in treating cancer in humans, these studies
A. display a high degree of statistical significance and so with nearly 100% certainty will work in humans.
B. are convincing, assuming the results have been replicated in a large number of mice.
C. are convincing, assuming that proper randomization and control were used.
D. suffer from lack of realism.
E. suffer from placebo effect.

___ 182. A lurking variable is
A. a variable that is not among the variables studied but that affects the response variable.
B. the true cause of a response.
C. any variable that produces a large residual.
D. the true variable that is explained by the explanatory variable.
E. another response variable.

___ 183. The owner of a chain of supermarkets notices that there is a positive correlation between the sales of beer and the sales of ice cream over the course of the previous year. During seasons when sales of beer were above average, sales of ice cream also tended to be above average. Likewise, during seasons when sales of beer were below average, sales of ice cream also tended to be below average. Which of the following would be a valid conclusion from these facts?
A. Sales records must be in error. There should be no association between beer and ice
cream sales.
B. Evidently, for a significant proportion of customers of these supermarkets, drinking beer causes a desire for ice cream or eating ice cream causes a thirst for beer.
C. A scatterplot of monthly ice cream sales versus monthly beer sales would show that a straight line describes the pattern in the plot, but it would have to be a horizontal line.
D. It is likely that sales of both beer and ice cream are confounded with a lurking variable, such as seasonal variation in temperature.
E. There is a clear, negative association between beer sales and ice cream sales.

184. A study of elementary school children, ages 6 to 11, finds a high positive correlation between shoe size $x$ and score $y$ on a test of reading comprehension. The observed correlation is most likely due to
A. the effect of a lurking variable, such as age.
B. a mistake, since the correlation must be negative.
C. cause and effect (larger shoe size causes higher reading comprehension).
D. "reverse" cause and effect (higher reading comprehension causes larger shoe size).
E. several outliers in the data set.

185. If changes in a response variable are due to the effects of the explanatory variable as well as the effects of lurking variables, and we cannot distinguish between these effects, we are said to have
A. a cause-and-effect relation between the explanatory and response variable.
B. a placebo effect.
C. confounding.
D. correlation.
E. extrapolated.

186. For one kindergarten class in his district, a researcher determines which children already can read simple words and which children cannot upon entering kindergarten. The children are followed until third grade, at which point they are tested to determine the grade level at which they are reading. Those children who were reading simple words on entering kindergarten are found to be reading at a higher level than those who could not read simple words on entering kindergarten. The researcher
A. can conclude that children should be taught to read in preschool, as there are clear benefits to reading early.
B. cannot conclude that being able to read before entering kindergarten is beneficial, as there may be confounding variables in this study.
C. needs to have taken a random sample of kindergarten students instead of one class to conclude a cause-and-effect relationship.
D. needs to check the reading level of the children's parents.
E. needs to retest in sixth grade or no conclusions can be reached.

187. The principle reason for the use of random assignment in designing experiments is that it
A. distinguishes a treatment effect from the effects of confounding variables.
B. allows double-blinding.
C. reduces sampling variability.
D. creates approximately equal groups for comparison.
E. eliminates the placebo effect.

188. The principle reason for the use of controls in designing experiments is that it
A. distinguishes a treatment effect from the effects of confounding variables.
B. allows double-blinding.
C. reduces sampling variability.
D. creates approximately equal groups for comparison.
E. eliminates the placebo effect.
189. The principle reason for replication in designing experiments is that it
A. distinguishes a treatment effect from the effects of confounding variables.
B. allows double-blinding.
C. reduces sampling variability.
D. creates approximately equal groups for comparison.
E. eliminates the placebo effect.

190. When controlled experiments are impractical or unethical, which of the following would be necessary to establish a cause-and-effect relation between two variables?
A. Strong association between the variables.
B. An association between the variables is observed in many different settings.
C. The alleged cause is plausible.
D. There is no obvious lurking variable that would affect the response variable.
E. All of the above.

191. In an experiment, an observed effect so large that it would rarely occur by chance is called
A. an outlier.
B. influential.
C. statistically significant.
D. bias.
E. replication.

192. One hundred volunteers who suffer from severe depression are available for a study. Fifty are selected at random and are given a new drug that is thought to be particularly effective in treating severe depression. The other fifty are given an existing drug for treating severe depression. A psychiatrist evaluates the symptoms of all volunteers after four weeks in order to determine if there has been substantial improvement in the severity of the depression. The study would be double blind if
A. neither drug had any identifying marks on it.
B. all volunteers were not allowed to see the psychiatrist nor the psychiatrist allowed to see the volunteers during the session in which the psychiatrist evaluated the severity of the depression.
C. neither the volunteers nor the psychiatrist knew which treatment any person had received.
D. the patients were given a placebo.
E. all of the above.

193. A double-blind experiment was conducted to evaluate the effectiveness of the Salk polio vaccine. The purpose of keeping the diagnosing physicians ignorant of the treatment status of the experimental subjects was to
A. eliminate grounds for malpractice suits.
B. ensure that subjects were randomly assigned to treatments.
C. eliminate a possible source of bias.
D. make sure nobody is harmed.
E. prevent stratification of the experiment.

194. In comparative trials in medicine, the placebo effect and subconscious bias on the part of the physicians evaluating treatment outcomes can be avoided by using
A. the double-blind technique.
B. randomized complete block designs.
C. response variables.
D. stratified random samples.
E. all of the above.
195. Twelve people who suffer from chronic fatigue syndrome volunteer to take part in an experiment to see if shark fin extract will increase one's energy level. Eight of the volunteers are men, and four are women. Half of the volunteers are to be given shark fin extract twice a day, and the other half are to be given a placebo twice a day. We wish to make sure that four men and two women are assigned to each of the treatments, so we decide to use a block design with the men forming one block and the women the other. A block design is appropriate in this experiment if
A. we want to be able to compare effects on energy level in men and women.
B. we believe men and women will respond differently to treatments.
C. gender equity is an important legal consideration in this study.
D. we want the conclusions to apply equally to men and women.
E. all of the above.

196. One hundred volunteers who suffer from severe depression are available for a study. Fifty are selected at random and are given a new drug that is thought to be particularly effective in treating severe depression. The other fifty are given an existing drug for treating severe depression. A psychiatrist evaluates the symptoms of all volunteers after four weeks in order to determine if there has been substantial improvement in the severity of the depression. The factor in this study is
A. which treatment the volunteers receive.
B. the extent to which the depression was reduced.
C. the use of randomization and the fact that this was a comparative study.
D. the use of a psychiatrist to evaluate the severity of depression.
E. the symptoms observed by the psychiatrist.

197. The reason that blocking (as in a randomized block design) is sometimes used in experimentation is to
A. prevent the placebo effect.
B. allow double blinding.
C. eliminate lurking variables
D. eliminate the need for random assignment.
E. reduce variability arising from random assignment.

198. Which of the following statements about a randomized complete block design with two treatments is not true?
A. Every experimental unit has a 50/50 chance of being given the first treatment.
B. Block A is chosen randomly from among the available experimental units.
C. In every block, some units are assigned the first treatment and some the second treatment.
D. Treatments are assigned randomly within each block.
E. All the units in Block A share the same (or similar) values of the blocked variable.

199. An experiment compares the taste of a new spaghetti sauce with the taste of a commercially successful sauce readily available in grocery stores. Each of a number of tasters tastes both sauces (in random order) and says which tastes better. This is called a
A. simple random sample.
B. stratified random sample.
C. completely randomized design.
D. matched pairs design.
E. double-blind design.

200. Are dogs better at tracking the movements of brightly colored objects? Fifteen experienced “disk dogs” who have been trained to catch flying disks in mid-air are given the chance to catch a bright red disk or a plain white disk. Each disk is thrown 10 times for each dog, with the sequence of disks (red or white) determined randomly. The proportion of red disks caught to the proportion of white disks caught is compared for each dog. This is an example of a
A. simple random sample.
B. stratified random sample.
C. completely randomized design.
D. matched pairs design.
E. double-blind design.

____ 201. Suppose we measure a response variable $Y$ for several values of an explanatory variable $X$. A scatterplot of log $Y$ versus log $X$ looks approximately like a negatively-sloping straight line. We may conclude that
A. the rate of growth of $Y$ is positive, but slowing down over time.
B. an exponential growth model would approximately describe the relationship between $Y$ and $X$.
C. a power model would approximately describe the relationship between $Y$ and $X$.
D. the relationship between $Y$ and $X$ is a positively-sloping straight line.
E. the residual plot of the regression of log $Y$ on log $X$ would have a “U-shaped” pattern suggesting a non-linear relationship.

____ 202. Using least-squares regression on data from 1990 through 2009, I determine that the (base 10) logarithm of the population of a country is approximately described by the equation 
$$\log \text{Population} = -13.5 + 0.01(\text{Year})$$
Which of the following is the predicted population of the country in the year 2010?
A. 6.6
B. 735
C. 2,000,000
D. 3,981,072
E. 33,000,000

____ 203. Which of the following would provide evidence that a power law model describes the relationship between a response variable $y$ and an explanatory variable $x$?
A. A normal probability plot of the residuals of the regression of log $y$ versus log $x$ looks approximately linear.
B. A normal probability plot of the residuals of the regression of log $y$ versus $x$ looks approximately linear.
C. A scatterplot of log $y$ versus $x$ looks approximately linear.
D. A scatterplot of $y$ versus log $x$ looks approximately linear.
E. A scatterplot of log $y$ versus log $x$ looks approximately linear.

____ 204. Suppose the relationship between a response variable $y$ and an explanatory variable $x$ is modeled well by the equation 
$$y = 3.6(0.32)^x$$
Which of the following plots is most likely to be roughly linear?
A. A plot of $y$ against $x$.
B. A plot of $y$ against log $x$.
C. A plot of log $y$ against $x$.
D. A plot of $10^y$ against $x$.
E. A plot of log $y$ against log $x$.

Scenario 12-7
Like most animals, small marine crustaceans are not able to digest all the food they eat. Moreover, the percentage of food eaten that is assimilated (that is, digested) decreases as the amount of food eaten increases. A scatterplot of this relationship for a certain species of crustacean (at right) indicates that it is non-linear. However, a scatterplot of In Assimilation versus In Food Intake is strongly linear. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.3324</td>
<td>0.5218</td>
<td>12.14</td>
<td>0.000</td>
</tr>
<tr>
<td>ln Food Intake</td>
<td>-0.6513</td>
<td>0.1047</td>
<td>-6.22</td>
<td>0.000</td>
</tr>
</tbody>
</table>

$S = 0.247460$  
$R^2 = 84.7\%$  
$R^2(adj) = 82.5\%$

---

### 205. Use Scenario 12-7. Which of the following best describe the model that is given by this computer printout?

A. A power model with equation $\ln \text{Assimilation} = 6.3324 - 0.6513 (\ln \text{Food Intake})$

B. An exponential model with equation $\ln \text{Assimilation} = 6.3324 - 0.6513 (\ln \text{Food Intake})$

C. A power model with equation $\ln \text{Food Intake} = -0.6513 + 6.3324 (\ln \text{Assimilation})$

D. An exponential model with equation $\ln \text{Food Intake} = -0.6513 + 6.3324 (\ln \text{Assimilation})$

E. A power model with equation $\ln \text{Assimilation} = -0.6513 + 6.3324 (\ln \text{Food Intake})$

---

### 206. Use Scenario 12-7. If, as described above, the scatterplot of $\ln$ Assimilation versus $\ln$ Food Intake is strongly linear, which of the following best describes the residual plot for the regression of these two variables?

A. A roughly straight line.

B. A “U-shaped” pattern, with positive residuals for low and high values of food intake and negative residuals in between.

C. A “U-shaped” pattern, with negative residuals for low and high values of food intake and positive residuals in between.

D. A curved pattern similar to the scatterplot of the variables Food Intake and Assimilation before the logarithmic transformation.

E. A random scattering of points on either side of the line whose equation is residuals = 0.

---

**Scenario 12-8**
Use of the Internet worldwide increased steadily from 1990 to 2002. A scatterplot of this growth (at right) shows a strongly non-linear pattern. However, a scatterplot of ln Internet Users versus Year is much closer to linear. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-951.10</td>
<td>43.45</td>
<td>-21.89</td>
<td>0.000</td>
</tr>
<tr>
<td>Year</td>
<td>0.4785</td>
<td>0.02176</td>
<td>21.99</td>
<td>0.000</td>
</tr>
<tr>
<td>S</td>
<td>0.2516</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Sq</td>
<td>98.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Sq(adj)</td>
<td>98.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scatterplot of internet users vs. year

207. Use Scenario 12-8. Which of the following best describe the model that is given by this computer printout?
   A. A power model with equation \( \ln \text{Internet Users} = -951.10 + 0.4785(\ln \text{Year}) \)
   B. A power model with equation \( \text{Internet Users} = -951.10 + 0.4785(\ln \text{Year}) \)
   C. A power model with equation \( \ln \text{Internet Users} = -951.10 + 0.4785(\text{Year}) \)
   D. An exponential model with equation \( \ln \text{Internet Users} = -951.10 + 0.4785(\text{Year}) \)
   E. An exponential model with equation \( \ln \text{Internet Users} = -951.10 + 0.4785(\ln \text{Year}) \)

208. Use Scenario 12-8. If, as described above, the scatterplot of ln Internet Users versus Year is strongly linear, which of the following best describes the residual plot for the regression of these two variables?
   A. A “U-shaped” pattern, with positive residuals for early and late years and negative residuals in between.
   B. A “U-shaped” pattern, with negative residuals for early and late years and positive residuals in between.
   C. A random scattering of points on either side of the line whose equation is residuals = 0.
   D. A curved pattern similar to the scatterplot of the variables Internet users and Year before the logarithmic transformation.
   E. A roughly straight line.

Scenario 12-9
Like most animals, small marine crustaceans are not able to digest all the food they eat. Moreover, the percentage of food eaten that is assimilated (that is, digested) decreases as the amount of food eaten increases. A residual plot for the regression of Assimilation rate (as a percentage of food intake) on Food Intake (in $\mu g$/day) is shown below.

___ 209. Use Scenario 12-9. Suppose we use the regression whose residuals are shown here to predict assimilation when food intake is 50 $\mu g$/day. Which of the following best describes the accuracy of that prediction.
A. The prediction would probably be accurate.
B. The prediction would probably overestimate the true assimilation rate.
C. The prediction would probably underestimate the true assimilation rate.
D. We do not have enough information to determine the accuracy of the estimate.
E. Whether the prediction underestimated or overestimated would depend on sampling variability, so that each type of error is equally likely.

___ 210. Use Scenario 12-9. A scatterplot of ln Assimilation versus ln Food Intake is strongly linear, suggesting that a linear regression of these transformed variables may be more appropriate. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.3324</td>
<td>0.5218</td>
<td>12.14</td>
<td>0.000</td>
</tr>
<tr>
<td>ln Food Intake</td>
<td>-0.6513</td>
<td>0.1047</td>
<td>-6.22</td>
<td>0.000</td>
</tr>
<tr>
<td>S = 0.247460</td>
<td>R-Sq  = 84.7%</td>
<td>R-Sq(adj) = 82.5%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When food intake is 250 $\mu g$/day, what is the predicted assimilation rate from this model?
A. 2.7%
B. 15.4%
C. 27.4%
D. 34.3%
E. 54.4%

Scenario 12-10
Use of the internet worldwide increased steadily from 1990 to 2002. A residual plot for the regression of worldwide Internet Users (in millions) on Year is shown below.
211. Use Scenario 12-10. Suppose we use the regression whose residuals are shown here to predict the number of Internet users in 1991. Which of the following best describes the accuracy of that prediction?
A. The prediction would probably underestimate the true number of Internet users in 1991.
B. The prediction would probably overestimate the true number of Internet users in 1991.
C. The prediction would probably be accurate.
D. We do not have enough information determine the accuracy of the estimate.
E. Since the sample is subject to random variable, the estimate will underestimate and overestimate with about the same frequency.

212. Use Scenario 12-10. A scatterplot of ln Internet Users (in millions) versus Year is strongly linear, suggesting that a linear regression of this transformation may be more appropriate. Below is a computer regression analysis of the transformed data (note that natural logarithms are used).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-951.10</td>
<td>43.45</td>
<td>-21.89</td>
<td>0.000</td>
</tr>
<tr>
<td>Year</td>
<td>0.4785</td>
<td>0.02176</td>
<td>21.99</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S = 0.2516  R-Sq = 98.2%  R-Sq(adj) = 98.0%

What is the predicted number of Internet users (in millions) in 1991, based on this model?
A. 1.59
B. 4.46
C. 4.92
D. 38.90
E. 86.77
MULTIPLE CHOICE

1. ANS: C  
   PTS: 1  
   TOP: Categorical vs Quantitative variables

2. ANS: B  
   PTS: 1  
   TOP: Categorical vs Quantitative variables

3. ANS: A  
   PTS: 1  
   TOP: Categorical vs Quantitative variables

4. ANS: C  
   PTS: 1  
   TOP: Categorical vs Quantitative variables

5. ANS: E  
   PTS: 1  
   TOP: Individuals and variables

6. ANS: C  
   PTS: 1  
   TOP: Individuals and variables

7. ANS: B  
   PTS: 1  
   TOP: Definition of distribution

8. ANS: C  
   PTS: 1  
   TOP: Definition of distribution

9. ANS: B  
   PTS: 1  
   TOP: Frequency table to pie chart

10. ANS: C  
    PTS: 1  
    TOP: When to use two-way tables

11. ANS: A  
    PTS: 1  
    TOP: Compare two categorical variables (not in two-way table)

12. ANS: D  
    PTS: 1  
    TOP: Conditional distribution—calculation

13. ANS: D  
    PTS: 1  
    TOP: Conditional distribution—identification

14. ANS: C  
    PTS: 1  
    TOP: Conditional distribution—calculation

15. ANS: C  
    PTS: 1  
    TOP: Conditional distribution—identification

16. ANS: E  
    PTS: 1  
    TOP: Segmented bar graphs

17. ANS: C  
    PTS: 1  
    TOP: Marginal distribution—calculation

18. ANS: B  
    PTS: 1  
    TOP: Conditional distribution—calculation

19. ANS: C  
    PTS: 1  
    TOP: Marginal distribution—identification

20. ANS: A  
    PTS: 1  
    TOP: Interpret two-way table

21. ANS: E  
    PTS: 1  
    TOP: Bar graph (including distortion)

22. ANS: C  
    PTS: 1  
    TOP: When to use bar graphs and pie charts

23. ANS: C  
    PTS: 1  
    TOP: Distorted pictogram

24. ANS: C  
    PTS: 1  
    TOP: Describing distribution; interpreting histogram

25. ANS: B  
    PTS: 1  
    TOP: Interpret histogram

26. ANS: D  
    PTS: 1  
    TOP: Interpret histogram

27. ANS: D  
    PTS: 1  
    TOP: Interpret histogram

28. ANS: D  
    PTS: 1  
    TOP: Interpret histogram

29. ANS: D  
    PTS: 1  
    TOP: Describing distribution; interpreting histogram

30. ANS: A  
    PTS: 1  
    TOP: Interpret stem plot

31. ANS: D  
    PTS: 1  
    TOP: Interpret stem plot

32. ANS: B  
    PTS: 1  
    TOP: Interpret stem plot

33. ANS: D  
    PTS: 1  
    TOP: Median and quartiles

34. ANS: C  
    PTS: 1  
    TOP: Histogram principles

35. ANS: B  
    PTS: 1  
    TOP: Calculating the mean

36. ANS: A  
    PTS: 1  
    TOP: Finding median from frequency distribution

37. ANS: E  
    PTS: 1  
    TOP: Behavior of median

38. ANS: C  
    PTS: 1  
    TOP: Mean, median, and skew

39. ANS: D  
    PTS: 1  
    TOP: Interpret segmented bar graph

40. ANS: B  
    PTS: 1  
    TOP: When to use segmented bar graph

41. ANS: B  
    PTS: 1  
    TOP: Median from dot plot
42. ANS: D
   PTS: 1
   TOP: IQR from dot plot
43. ANS: B
   PTS: 1
   TOP: Box plot from dot plot
44. ANS: B
   PTS: 1
   TOP: Calculating the mean
45. ANS: C
   PTS: 1
   TOP: Interpret boxplot
46. ANS: A
   PTS: 1
   TOP: Interpret boxplot
47. ANS: D
   PTS: 1
   TOP: Interpret boxplot
48. ANS: B
   PTS: 1
   TOP: Interpret boxplot
49. ANS: D
   PTS: 1
   TOP: Interpret boxplot
50. ANS: B
   PTS: 1
   TOP: Mean and standard deviation behavior
51. ANS: B
   PTS: 1
   TOP: Standard deviation and variance relationship
52. ANS: C
   PTS: 1
   TOP: Standard deviation and variance units
53. ANS: D
   PTS: 1
   TOP: Calculating variance
54. ANS: E
   PTS: 1
   TOP: Units for numerical measures
55. ANS: B
   PTS: 1
   TOP: Choosing the right measures
56. ANS: E
   PTS: 1
   TOP: Choosing the right measures
57. ANS: B
   PTS: 1
   TOP: 1.5 x IQR rule (from data)
58. ANS: A
   PTS: 1
   TOP: 1.5 x IQR rule (from 5-number summary)
59. ANS: D
   PTS: 1
   TOP: Compare parallel box plots
60. ANS: C
   PTS: 1
   TOP: Interpret stem plot
61. ANS: B
   PTS: 1
   TOP: Simpson's paradox
62. ANS: C
   PTS: 1
   TOP: Simpson's paradox
63. ANS: D
   PTS: 1
   TOP: Simpson's paradox
64. ANS: B
   PTS: 1
   TOP: Explanatory/response
65. ANS: E
   PTS: 1
   TOP: Interpreting Scatterplot
66. ANS: D
   PTS: 1
   TOP: Impact of Outlier on r
67. ANS: A
   PTS: 1
   TOP: Explanatory/response
68. ANS: B
   PTS: 1
   TOP: Explanatory/response
69. ANS: B
   PTS: 1
   TOP: Negative association
70. ANS: E
   PTS: 1
   TOP: Scatterplot basics
71. ANS: D
   PTS: 1
   TOP: Scatterplot basics
72. ANS: C
   PTS: 1
   TOP: Interpreting scatterplot
73. ANS: B
   PTS: 1
   TOP: Scatterplot basics
74. ANS: B
   PTS: 1
   TOP: Causation
75. ANS: A
   PTS: 1
   TOP: Estimating r from scatter
76. ANS: C
   PTS: 1
   TOP: Characteristics of r—changing units
77. ANS: E
   PTS: 1
   TOP: Characteristics of r—unitless measure
78. ANS: C
   PTS: 1
   TOP: Causation
79. ANS: D
   PTS: 1
   TOP: Scatterplot with categorical variable
80. ANS: E
   PTS: 1
   TOP: Interpreting correlation
81. ANS: E
   PTS: 1
   TOP: Interpreting correlation
82. ANS: B
   PTS: 1
   TOP: Estimating r from scatter
83. ANS: A
   PTS: 1
   TOP: Impact of outlier on r
84. ANS: C
   PTS: 1
   TOP: Characteristics of r—changing units
85. ANS: B
   PTS: 1
   TOP: Characteristics of r—several
86. ANS: C
   PTS: 1
   TOP: Non-linear data and r
87. ANS: D
   PTS: 1
   TOP: Interpreting correlation
88. ANS: B
   PTS: 1
   TOP: How r is calculated
89. ANS: A  PTS: 1  TOP: How r is calculated
90. ANS: D  PTS: 1  TOP: Scatterplot with categorical variable
91. ANS: C  PTS: 1  TOP: Regression line basics (y-intercept)
92. ANS: D  PTS: 1  TOP: Prediction estimate from graph
93. ANS: A  PTS: 1  TOP: Residual estimate from graph
94. ANS: C  PTS: 1  TOP: Explanatory/response
95. ANS: D  PTS: 1  TOP: Prediction estimate from graph
96. ANS: A  PTS: 1  TOP: What least-squares means
97. ANS: C  PTS: 1  TOP: Y-intercept from formula
98. ANS: D  PTS: 1  TOP: Interpreting residual
99. ANS: E  PTS: 1  TOP: Interpret residuals
100. ANS: A  PTS: 1  TOP: Interpret r-sq
101. ANS: B  PTS: 1  TOP: Interpret r-sq
102. ANS: D  PTS: 1  TOP: Mean of residuals = 0
103. ANS: A  PTS: 1  TOP: Interpreting residual plot
104. ANS: C  PTS: 1  TOP: Influential points and outliers
105. ANS: C  PTS: 1  TOP: Influential points and outliers
106. ANS: B  PTS: 1  TOP: Extrapolation
107. ANS: B  PTS: 1  TOP: Interpret slope/computer output
108. ANS: D  PTS: 1  TOP: Interpret s from computer output
109. ANS: B  PTS: 1  TOP: Calculate residual
110. ANS: E  PTS: 1  TOP: Extrapolation
111. ANS: B  PTS: 1  TOP: Interpret residual
112. ANS: E  PTS: 1  TOP: Interpret residual
113. ANS: B  PTS: 1  TOP: Interpret slope
114. ANS: A  PTS: 1  TOP: Interpret computer output
115. ANS: A  PTS: 1  TOP: Identify s from computer output
116. ANS: B  PTS: 1  TOP: Interpret r-sq
117. ANS: D  PTS: 1  TOP: Interpret residuals
118. ANS: D  PTS: 1  TOP: Interpret r-sq
119. ANS: D  PTS: 1  TOP: Prediction
120. ANS: D  PTS: 1  TOP: Interpret slope
121. ANS: B  PTS: 1  TOP: Interpret y-intercept
122. ANS: B  PTS: 1  TOP: Interpret r-sq
123. ANS: D  PTS: 1  TOP: Regression slope from formula
124. ANS: C  PTS: 1  TOP: Characteristics of LSRL
125. ANS: E  PTS: 1  TOP: Interpret s
126. ANS: A  PTS: 1  TOP: Identify population
127. ANS: D  PTS: 1  TOP: Identify sample
128. ANS: D  PTS: 1  TOP: Convenience sample
129. ANS: C  PTS: 1  TOP: Identify sampling frame
130. ANS: A  PTS: 1  TOP: Identify sample
131. ANS: D  PTS: 1  TOP: Bias from poor sampling frame
132. ANS: A  PTS: 1  TOP: Sample and sampling frame
133. ANS: D  PTS: 1  TOP: Convenience sample
134. ANS: A  PTS: 1  TOP: Voluntary response
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Points</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>135.</td>
<td>A</td>
<td>1</td>
<td>Voluntary response</td>
</tr>
<tr>
<td>136.</td>
<td>A</td>
<td>1</td>
<td>Voluntary response</td>
</tr>
<tr>
<td>137.</td>
<td>C</td>
<td>1</td>
<td>SRS definition</td>
</tr>
<tr>
<td>138.</td>
<td>A</td>
<td>1</td>
<td>SRS definition</td>
</tr>
<tr>
<td>139.</td>
<td>C</td>
<td>1</td>
<td>SRS definition</td>
</tr>
<tr>
<td>140.</td>
<td>E</td>
<td>1</td>
<td>What a SRS doesn't do</td>
</tr>
<tr>
<td>141.</td>
<td>E</td>
<td>1</td>
<td>Non-sampling error</td>
</tr>
<tr>
<td>142.</td>
<td>C</td>
<td>1</td>
<td>Using a random digits table</td>
</tr>
<tr>
<td>143.</td>
<td>A</td>
<td>1</td>
<td>Idea of random digits table</td>
</tr>
<tr>
<td>144.</td>
<td>E</td>
<td>1</td>
<td>Idea of random digits table</td>
</tr>
<tr>
<td>145.</td>
<td>C</td>
<td>1</td>
<td>Using a random digits table</td>
</tr>
<tr>
<td>146.</td>
<td>E</td>
<td>1</td>
<td>Using a random digits table</td>
</tr>
<tr>
<td>147.</td>
<td>D</td>
<td>1</td>
<td>Using a random digits table</td>
</tr>
<tr>
<td>148.</td>
<td>B</td>
<td>1</td>
<td>Stratified random sample</td>
</tr>
<tr>
<td>149.</td>
<td>C</td>
<td>1</td>
<td>Why stratify</td>
</tr>
<tr>
<td>150.</td>
<td>A</td>
<td>1</td>
<td>Stratification and blocking</td>
</tr>
<tr>
<td>151.</td>
<td>D</td>
<td>1</td>
<td>What kind of sampling? (Stratified)</td>
</tr>
<tr>
<td>152.</td>
<td>B</td>
<td>1</td>
<td>What kind of sampling? (Cluster)</td>
</tr>
<tr>
<td>153.</td>
<td>C</td>
<td>1</td>
<td>What kind of sampling? (Multistage)</td>
</tr>
<tr>
<td>154.</td>
<td>E</td>
<td>1</td>
<td>What kind of sampling? (Multistage)</td>
</tr>
<tr>
<td>155.</td>
<td>E</td>
<td>1</td>
<td>What kind of sampling? (Stratified)</td>
</tr>
<tr>
<td>156.</td>
<td>B</td>
<td>1</td>
<td>What kind of sampling? (Cluster)</td>
</tr>
<tr>
<td>157.</td>
<td>B</td>
<td>1</td>
<td>Question wording</td>
</tr>
<tr>
<td>158.</td>
<td>B</td>
<td>1</td>
<td>Question wording</td>
</tr>
<tr>
<td>159.</td>
<td>B</td>
<td>1</td>
<td>Question wording</td>
</tr>
<tr>
<td>160.</td>
<td>D</td>
<td>1</td>
<td>What kind of bias? (Nonresponse)</td>
</tr>
<tr>
<td>161.</td>
<td>E</td>
<td>1</td>
<td>What kind of bias? (Response bias)</td>
</tr>
<tr>
<td>162.</td>
<td>C</td>
<td>1</td>
<td>What kind of bias? (Undercoverage)</td>
</tr>
<tr>
<td>163.</td>
<td>C</td>
<td>1</td>
<td>Principles of experimental design</td>
</tr>
<tr>
<td>164.</td>
<td>E</td>
<td>1</td>
<td>Experiment vs. Observational study</td>
</tr>
<tr>
<td>165.</td>
<td>A</td>
<td>1</td>
<td>Experiment vs. Observational study</td>
</tr>
<tr>
<td>166.</td>
<td>B</td>
<td>1</td>
<td>Scope of inference</td>
</tr>
<tr>
<td>167.</td>
<td>B</td>
<td>1</td>
<td>Identify response variable</td>
</tr>
<tr>
<td>168.</td>
<td>A</td>
<td>1</td>
<td>Experiment vs. Observational study</td>
</tr>
<tr>
<td>169.</td>
<td>C</td>
<td>1</td>
<td>Confounding</td>
</tr>
<tr>
<td>170.</td>
<td>E</td>
<td>1</td>
<td>Identify response variable</td>
</tr>
<tr>
<td>171.</td>
<td>A</td>
<td>1</td>
<td>Placebo effect</td>
</tr>
<tr>
<td>172.</td>
<td>E</td>
<td>1</td>
<td>Block design</td>
</tr>
<tr>
<td>173.</td>
<td>D</td>
<td>1</td>
<td>Experiment vs. Observational study</td>
</tr>
<tr>
<td>174.</td>
<td>C</td>
<td>1</td>
<td>Identify explanatory variable</td>
</tr>
<tr>
<td>175.</td>
<td>A</td>
<td>1</td>
<td>Confounding</td>
</tr>
<tr>
<td>176.</td>
<td>E</td>
<td>1</td>
<td>Block design</td>
</tr>
<tr>
<td>177.</td>
<td>B</td>
<td>1</td>
<td>Identify response variable</td>
</tr>
<tr>
<td>178.</td>
<td>A</td>
<td>1</td>
<td>Identify factors</td>
</tr>
<tr>
<td>179.</td>
<td>B</td>
<td>1</td>
<td>Identify experimental units</td>
</tr>
<tr>
<td>180.</td>
<td>A</td>
<td>1</td>
<td>Scope of inference</td>
</tr>
<tr>
<td>181.</td>
<td>D</td>
<td>1</td>
<td>Lack of realism</td>
</tr>
</tbody>
</table>
182. ANS: A  PTS: 1  TOP: Lurking variable
183. ANS: D  PTS: 1  TOP: Confounding
184. ANS: A  PTS: 1  TOP: Confounding
185. ANS: C  PTS: 1  TOP: Confounding
186. ANS: B  PTS: 1  TOP: Confounding
187. ANS: D  PTS: 1  TOP: Purpose of randomization
188. ANS: A  PTS: 1  TOP: Purpose of control
189. ANS: C  PTS: 1  TOP: Purpose of replication
190. ANS: E  PTS: 1  TOP: Establishing cause and effect
191. ANS: C  PTS: 1  TOP: Statistical significance
192. ANS: C  PTS: 1  TOP: Double blind
193. ANS: C  PTS: 1  TOP: Double blind
194. ANS: A  PTS: 1  TOP: Double blind
195. ANS: B  PTS: 1  TOP: Block design
196. ANS: A  PTS: 1  TOP: Identify factor
197. ANS: E  PTS: 1  TOP: Block design
198. ANS: B  PTS: 1  TOP: Block design
199. ANS: D  PTS: 1  TOP: Matched Pairs
200. ANS: D  PTS: 1  TOP: Matched Pairs
201. ANS: C  PTS: 1  TOP: Interpreting log-log scatterplot
202. ANS: D  PTS: 1  TOP: Prediction from log Y vs. X regression
203. ANS: E  PTS: 1  TOP: Power functions and transformations
204. ANS: C  PTS: 1  TOP: Exponential functions and transformations
205. ANS: A  PTS: 1  TOP: Model/Equation for log-log transformation
206. ANS: E  PTS: 1  TOP: Residuals for good linear fit
207. ANS: D  PTS: 1  TOP: Model/Equation for semi-log transformation
208. ANS: C  PTS: 1  TOP: Residuals for good linear fit
209. ANS: C  PTS: 1  TOP: Accuracy of estimate from residuals
210. ANS: B  PTS: 1  TOP: Prediction from log Y vs. log X regression
211. ANS: A  PTS: 1  TOP: Accuracy of estimate from residuals
212. ANS: C  PTS: 1  TOP: Prediction from log Y vs. X regression