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 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

3. 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.

4. 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.

5. 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. ![Pie Chart A]
B. ![Pie Chart B]
C. ![Pie Chart C]
D. ![Pie Chart D]
E. ![Pie Chart E]
6. 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></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>

7. 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.

8. 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.

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

10. Use Scenario 1-1. 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.

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

Where does the 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.
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.

13. 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.

14. 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%.

15. 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.
16. 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, 46, 79
D. 116, 118, 121, 124, 128, 133, 137, 142, 146, 179
E. None of the above.

17. 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.

18. 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
C. $15
D. $20
E. $25

19. 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.

20. 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.

21. Use Scenario 1-4. 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.
22. Use Scenario 1-4. Which of the following is a correct box plot for these data?

A. A  
B. B  
C. C  
D. D  
E. E

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.

23. 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.

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.

24. 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.

25. 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.

26. 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.

27. 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.
28. 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.

29. 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

30. 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.

31. For the density curve below, which of the following is true?
A. The median is 0.5.
B. The median is larger than 0.5.
C. The density curve is skewed right.
D. The density curve is Normal.
E. The density curve is symmetric.
32. Use Figure 1-1. For this density curve, what percent of the observations lie above 4?
   A. 20%
   B. 25%
   C. 50%
   D. 75%
   E. 80%

33. You can roughly locate the mean of a density curve by eye because it is
   A. the point at which the curve would balance if made of solid material.
   B. the point that divides the area under the curve into two equal parts.
   C. the point at which the curve reaches its peak.
   D. the point where the curvature changes direction.
   E. the point at which the height of the graph is equal to 1.

34. The density curve below shows the distribution of a variable that takes values from 0 to 1. What percent of the observations lie between 0 and 0.5?
   A. 20%
   B. 25%
   C. 50%
   D. 75%
   E. 80%

35. The 35th percentile of a population is the number x such that
   A. 35% of the population scores are above x.
   B. 65% of the population scores are above x.
   C. 35% of the population scores equal x.
   D. x is the 35th percentile.
   E. x is 35% of the population mean.

36. A score’s percentile is a measure of
   A. center
   B. spread
   C. relative location
   D. skew
   E. relative frequency

37. Consider the following cumulative relative frequency graph of the scores of students in an introductory statistics course:

   A grade of C or C+ is assigned to a student who scores between 55 and 70. The percentage of students who obtained a grade of C or C+ is
   A. 15%
   B. 20%
   C. 25%
   D. 30%
   E. 50%

38. A standard score describes
   A. How much skew there is in a distribution
   B. How much spread there is in a distribution
   C. How far apart the mean and median of a distribution are.
   D. How far a particular score is from the mean.
   E. How far a particular score is from the median.

39. The mean number of days that the midge Chaoborus spends in its larval stage is 14.1 days, with a standard deviation of 2.2 days. This distribution is skewed toward higher values. What is the z-score for an individual midge that spends 12.7 days in its larval stage?
   A. –1.11
   B. –0.64
   C. 0.64
   D. 0.94
   E. None of these, because z-score cannot be used unless the distribution is Normal.

40. Ramon is planning on buying a new car. He’s looking at the Ford Escape—a sport-utility vehicle—which gets 28 highway miles per gallon, and the Ford Fusion—a mid-sized sedan—which gets 31 highway miles per gallon. The mean fuel efficiency for all sport utility vehicles is 23, with a standard deviation of 7.6. The mean of all mid-sized sedans is 27, with a standard deviation of 5.2. Which vehicle has a better standing, relative to others of the same style?
   A. The Ford Fusion sedan has a better relative standing, because it’s z-score is higher.
   B. The Ford Fusion sedan has a better relative standing, because it’s z-score is closer to 0.
   C. The Ford Escape SUV has a better relative standing, because it’s z-score is closer to 0.
   D. The Ford Escape SUV has a better relative standing, because it’s z-score is higher.
   E. We can’t make any comparisons unless we know that the distribution of fuel efficiency for vehicle types is Normally distributed.

41. If 30 is added to every number on a list, the only one of the following that is not changed is
   A. the mean.
   B. the mode.
   C. the 75th percentile.
   D. the median.
   E. the standard deviation.
Scenario 2-1
A sample was taken of the salaries of 20 employees of a large company. The following are the salaries (in thousands of dollars) for this year. For convenience, the data are ordered.

<table>
<thead>
<tr>
<th>Salary (in thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>37</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>49</td>
</tr>
<tr>
<td>51</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>61</td>
</tr>
<tr>
<td>67</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>77</td>
</tr>
</tbody>
</table>

Suppose each employee in the company receives a $3,000 raise for next year (each employee's salary is increased by $3,000).

Use Scenario 2-1. The median salary for the employees working for the company will
A. be unchanged.
B. increase by $3,000.
C. be multiplied by $3,000.
D. increase by $3,000
E. increase by $150.

Use Scenario 2-1. The interquartile range of the salaries for the employees will
A. be unchanged.
B. increase by $3,000.
C. be multiplied by $3,000.
D. increase by $3,000
E. decrease by $3,000.

The Normal curve below describes the death rates per 100,000 people in developed countries in the 1990's.

The mean and standard deviation of this distribution are approximately
A. Mean ≈ 100; Standard Deviation ≈ 65
B. Mean ≈ 100; Standard Deviation ≈ 100
C. Mean ≈ 150; Standard Deviation ≈ 65
D. Mean ≈ 190; Standard Deviation ≈ 100
E. Mean ≈ 200; Standard Deviation ≈ 130

48. Using the standard Normal distribution tables, the area under the standard Normal curve corresponding to \( Z < 1.1 \) is
A. 0.1557.
B. 0.2704.
C. 0.8413.
D. 0.8478.
E. 0.8643.

49. Using the standard Normal distribution tables, the area under the standard Normal curve corresponding to \(-0.5 < Z < 1.2\) is
A. 0.2815.
B. 0.3085.
C. 0.3661.
D. 0.5764.
E. 0.8849.

50. The scores on a university examination are Normally distributed with a mean of 62 and a standard deviation of 11. If the bottom 5% of students will fail the course, what is the lowest mark that a student can have and still be awarded a passing grade?
A. 40
B. 43
C. 44
D. 57
E. 62

51. A company produces packets of soap powder labeled "Giant Size 32 Ounces." The actual weight of soap powder in a box has a Normal distribution with a mean of 33 oz. and a standard deviation of 0.8 oz. What proportion of packets are underweight (i.e., weigh less than 32 oz.)?
A. 0.106.
B. 0.115.
C. 0.159.
D. 0.212.
E. 0.841.

52. The time to complete a standardized exam is approximately Normal with a mean of 70 minutes and a standard deviation of 10 minutes. How much time should be given to complete the exam so that 80% of the students will complete the exam in the time given?
A. 61.6 minutes
B. 78.4 minutes
C. 79.8 minutes
D. 84 minutes
E. 92.8 minutes

53. Suppose that scores on a certain IQ test are Normally distributed with mean 110 and standard deviation 15. Then about 40% of the scores are between
A. 80 and 140.
B. 65 and 155.
C. 106 and 110.
D. the 25th and 75th percentiles.
E. 102 and 118.

54. You are chatting with the principal of a local high school. The topic of SAT scores comes up, and the principal mentions that SAT scores at the school are Normally distributed. She doesn't remember the mean or the standard deviation, but she does remember that the upper and lower quartiles are 500 and 600. The standard deviation of SAT verbal scores is closest to
A. 25 points.
B. 50 points.
C. 75 points.
D. 100 points.
E. 550 points.

55. A stemplot of a set of data is roughly symmetric, but the data do not even approximately follow the 68–95–99.7 rule. We conclude that the data are
A. Normal, but not Standard Normal.
B. Standard Normal.
C. not Normal.
D. Normal.
E. skewed in both directions.
56. The plot shown below is a Normal probability plot for the total annual cost (tuition plus room and board) to attend 126 of the top colleges in the country in 2005. Which statement is true for these data?

A. The data are clearly Normally distributed.
B. The data are approximately Normally distributed.
C. The data are clearly skewed to the left.
D. The data are clearly skewed to the right.
E. There is insufficient information to determine the shape of the distribution.

57. 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.

58. 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 its value does not depend which variable is used for x and which is used for y.

59. 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.

60. 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.
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.

<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 King</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dom</td>
<td>1.3</td>
<td>199</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>

62. 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.

63. 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.

64. 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.

65. 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.

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.

66. 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.
67. 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.

68. 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.

69. 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.

70. 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.

![Nitrogen oxide vs carbon monoxide](image)

71. 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

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 $\hat{Y} = 10 + 0.5X$, where $\hat{Y}$ represents the predicted final exam score and $X$ is the score on the first exam.

72. Use Scenario 3-5. The first test score is
   A. the intercept.
   B. the slope.
   C. the explanatory variable.
   D. the response variable.
   E. a lurking variable.

73. “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.

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.

$r = \text{correlation between } X \text{ and } Y = 0.9$

$J = \text{mean of the values of } X = 6.5$
$M = \text{mean of the values of } Y = 6.6$
$Sx = \text{standard deviation of the values of } X = 3.6$
$Sy = \text{standard deviation of the values of } Y = 1.2$

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

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).

82. 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)

83. 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.
84. 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: $\hat{y} = 1344 - 15\sqrt{x}$

85. 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.

86. 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. The correlation between Exam 1 scores and Exam 2 scores is $r = 0.6$.

87. 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. The correlation coefficient would change, but the regression line would not change.
B. Both the correlation coefficient and the regression line would be unchanged.
C. The correlation coefficient would not change, but the regression line would change.
D. The correlation coefficient would change but the regression equation would remain the same.
E. Only the $y$-intercept of the regression line would change, the slope of the line and the correlation coefficient would not change.

88. 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.

89. 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 underestimate the level of support among all Lafayette residents.
E. This is a convenience sample. It will almost certainly overestimate the level of support among all Lafayette residents.

Scenario 4-2
You want 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.

90. 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.

91. 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.

92. 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.

93. 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 currently on a diet. It gives very accurate results.

94. 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. The correlation between Exam 1 scores and Exam 2 scores is $r = 0.6$.
94. 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.

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 20498 05507 11391 46082 20751 27498 12909 45287 71753 98236 66419 84533

95. 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.

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.

9592 94007 69769 33547 72450 16632 81194 14873

96. 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.

97. 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.

98. 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.

99. 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.

100. 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.

101. To determine the proportion of each color of Peanut Butter M&M, you buy 10 1.69 ounce packages and measured the size of every barnacle in each plot. This is an example of
A. simple random sampling
B. cluster sampling
C. multistage sampling
D. stratified random sampling
E. systematic random sampling

102. 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.

103. 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.

104. 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.
105. 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. bias resulting from question wording. C. nonresponse. D. response bias. E. a convenience sample.

106. Just before the presidential election of 1936, the magazine Literary Digest predicted—in correctly, 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.

107. 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.

108. 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.

109. 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.

110. 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.

111. 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.

112. 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.

113. 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.

114. 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 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.

115. Use Scenario 4-8. Which of the following best describes the inferences the researchers can make based on 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.

116. 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.

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.

118. 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.

119. 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.

120. When controlled experiments are impractical or unethical, which of the following would be necessary to establish a cause-and-effect relationship 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.

121. 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.

122. 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.

123. 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.

124. 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.

125. 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.
126. I toss a penny and observe whether it lands heads up or tails up. Suppose the penny is fair, i.e., the probability of heads is 1/2 and the probability of tails is 1/2. This means that
A. every occurrence of a head must be balanced by a tail in one of the next two or three tosses.
B. if I flip the coin 10 times, it would be almost impossible to obtain 7 heads and 3 tails.
C. if I flip the coin many, many times the proportion of heads will be approximately 1/2, and this proportion will tend to get closer and closer to 1/2 as the number of tosses increases.
D. regardless of the number of flips, half will be heads and half tails.
E. all of the above.

127. When two coins are tossed, the probability of getting two heads is 0.25. This means that
A. of every 100 tosses, exactly 25 will have two heads.
B. the odds against two heads are 4 to 1.
C. in the long run, the average number of heads is 0.25.
D. in the long run two heads will occur on 25% of all tosses.
E. if you get two heads on each of the first five tosses of the coins, you are unlikely to get heads the fourth time.

128. You read in a book on poker that the probability of being dealt three of a kind in a five-card poker hand is 1/50. What does this mean?
A. 1
B. 2
C. 3
D. 4
E. 5

129. In probability and statistics, a random phenomenon is
A. something that is completely unexpected or surprising
B. something that has a limited set of outcomes, but when each outcome occurs is completely unpredictable.
C. something that appears unpredictable, but each individual outcome can be accurately predicted with appropriate mathematical or computer modeling.
D. something that is unpredictable from one occurrence to the next, but over the course of many occurrences follows a predictable pattern.
E. something whose outcome defies description.

130. A poker player is dealt poor hands for several hours. He decides to bet heavily on the last hand of the evening on the grounds that after many bad hands he is due for a winner.
A. He's right, because the winnings have to average out.
B. He's wrong, because successive deals are independent of each other.
C. He's right, because successive deals are independent of each other.
D. He's wrong, because he's clearly on a "cold streak."
E. Whether he's right or wrong depends on how many bad hands he's been dealt so far.

131. A box has 10 tickets in it, two of which are winning tickets. You draw a ticket at random. If it's a winning ticket, you win. If not, you get another chance, as follows: your losing ticket is replaced in the box by a winning ticket (so now there are 10 tickets, as before, but 3 of them are winning tickets). You get to draw again, at random. Which of the following are legitimate methods for using simulation to estimate the probability of winning?
I. Choose, at random, a two-digit number. If the first digit is 0 or 1, you win on the first draw. If the first digit is 2 through 9, but the second digit is 0, 1, or 2, you win on the second draw. Any other two-digit number means you lose.
II. Choose, at random, a one-digit number. If it is 0 or 1, you win. If it is 2 through 9, pick a second number. If the second number is 8, 9, or 0, you win. Otherwise, you lose.
III. Choose, at random, a one-digit number. If it is 0 or 1, you win on the first draw. If it is 2, 3, or 4, you win on the second draw; If it is 5 through 9, you lose.
A. I only
B. II only
C. III only
D. I and II
E. I, II, and III

132. A basketball player makes 75% of his free throws. We want to estimate the probability that he makes 4 or more frees throws out of 5 attempts (we assume the shots are independent). To do this, we use the digits 1, 2, and 3 to correspond to making the free throw and the digit 4 to correspond to missing the free throw. If the table of random digits begins with the digits below, how many free throw does he hit
A. 1
B. 2
C. 3
D. 4
E. 5

133. Use Scenario 5-1. Based on your simulation, the estimated probability of winning nothing is
A. 1/2
B. 2/11
C. 2/11
D. 6/15
E. 7/11

134. I select two cards from a deck of 52 cards and observe the color of each (26 cards in the deck are red and 26 are black). Which of the following is an appropriate sample space S for the possible outcomes?
A. S = (red, black)
B. S = (red, red), (red, black), (black, red), (black, black)
C. something that appears unpredictable, but each individual outcome can be accurately predicted with appropriate mathematical or computer modeling.
D. something that is unpredictable from one occurrence to the next, but over the course of many occurrences follows a predictable pattern.
E. something whose outcome defies description.

135. A game consists of drawing three cards at random from a deck of playing cards. You win $3 for each red card that is drawn. It costs $2 to play. For one play of this game, the sample space S for the net amount you win (after deducting the cost of play) is
A. S = [$0, $1, $2, $3]
B. S = [$6, $7, $8, $9]
C. S = -[$2, $1, $4, $7]
D. S = -[$2, $3, $6, $9]
E. S = [$0, $3, $6, $9]
136. An assignment of probabilities must obey which of the following?
A. The probability of any event must be a number between 0 and 1, inclusive.
B. The sum of all the probabilities of all outcomes in the sample space must be exactly 1.
C. The probability of an event is the sum of the probabilities of outcomes in the sample space in which the event occurs.
D. All three of the above.
E. A and B only.

137. Event $A$ has probability 0.4. Event $B$ has probability 0.5. If $A$ and $B$ are independent, then the events are said to be
A. independent.
B. disjoint.
C. mutually exhaustive.
D. the sample space.
E. complementary.

Scenario 5-3
Ignoring twins and other multiple births, assume babies born at a hospital are independent random events with the probability that a baby is a boy and the probability that a baby is a girl both equal to 0.5.

138. Use Scenario 5-3. The probability that at least one of the next three babies is a boy is
A. 0.125.
B. 0.333.
C. 0.667.
D. 0.750.
E. 0.875.

139. Here is an assignment of probabilities to the face that comes up when rolling a die once:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>1/7</td>
<td>2/7</td>
<td>0</td>
<td>3/7</td>
<td>0</td>
<td>1/7</td>
</tr>
</tbody>
</table>

Which of the following is true?
A. This isn't a legitimate assignment of probability, because every face of a die must have probability 1/6.
B. This isn't a legitimate assignment of probability, because it gives probability zero to rolling a 3 or a 5.
C. This isn't a legitimate assignment of probability, because the probabilities do not add to exactly 1.
D. This isn't a legitimate assignment of probability, because we must actually roll the die many times to learn the true probabilities.
E. This is a legitimate assignment of probability.

137. Event $A$ has probability 0.4. Event $B$ has probability 0.5. If $A$ and $B$ are independent, then the events are said to be
A. independent.
B. disjoint.
C. mutually exhaustive.
D. the sample space.
E. complementary.

Scenario 5-3
Ignoring twins and other multiple births, assume babies born at a hospital are independent random events with the probability that a baby is a boy and the probability that a baby is a girl both equal to 0.5.

140. If the knowledge that an event $A$ has occurred implies that a second event $B$ cannot occur, the events are said to be
A. independent.
B. disjoint.
C. mutually exhaustive.
D. the sample space.
E. complementary.

Scenario 5-4
In a particular game, a fair die is tossed. If the number of spots showing is either four or five, you win $1. If the number of spots showing is six, you win $4. And if the number of spots showing is one, two, or three, you win nothing. You are going to play the game twice.

145. Use Scenario 5-4. The probability that you win $4 both times is
A. 1/36.
B. 1/12
C. 1/6.
D. 1/4.
E. 1/3.

Scenario 5-5
Suppose we roll two six-sided dice—one red and one green. Let $A$ be the event that the number of spots showing on the red die is three or less and $B$ be the event that the number of spots showing on the green die is three or more.

146. Use Scenario 5-5. The events $A$ and $B$ are
A. disjoint.
B. conditional.
C. independent.
D. reciprocals.
E. complementary.

147. Use Scenario 5-5. $P(A \cup B) =$
A. 1/6.
B. 1/4
C. 2/3.
D. 5/6.
E. 1.
Scenario 5-6
A system has two components that operate in parallel, as shown in the diagram below. Because the components operate in parallel, at least one of the components must function properly if the system is to function properly. Let $F$ denote the event that component 1 fails during one period of operation and $G$ denote the event that component 2 fails during one period of operation. Suppose $P(F) = 0.20$ and $P(G) = 0.03$. The component failures are independent.

148. Use Scenario 5-6. The event corresponding to the system functioning properly during one period of operation is
A. $F$ and $G$.
B. $F$ or $G$.
C. not $F$ or not $G$.
D. not $F$ and not $G$.
E. not $F$ or $G$.

149. Event $A$ occurs with probability 0.8. The conditional probability that event $B$ occurs, given that $A$ occurs, is 0.5. The probability that both $A$ and $B$ occur
A. is 0.1.
B. is 0.4.
C. is 0.625.
D. is 0.8.
E. cannot be determined from the information given.

150. The card game Euchre uses a deck with 32 cards: Ace, King, Queen, Jack, 10, 9, 8, 7 of each suit. Suppose you choose one card at random from a well-shuffled Euchre deck. What is the probability that the card is a Jack, given that you know it’s a face card?
A. 1/3
B. 1/4
C. 1/8
D. 1/9
E. 1/12

151. Among the students at a large university who describe themselves as vegetarians, some eat fish, some eat eggs, some eat both fish and eggs, and some eat neither fish nor eggs. Choose a vegetarian student at random. Let $E$ = the event that the student eats eggs, and let $F$ = the event that the student eats fish. Which of the following Venn diagrams has correctly shaded the event that the student eats neither fish nor eggs?

A. 
B. 
C. 
D. 
E. 

Scenario 5-7
The probability of a randomly selected adult having a rare disease for which a diagnostic test has been developed is 0.001. The diagnostic test is not perfect. The probability the test will be positive (indicating that the person has the disease) is 0.99 for a person with the disease and 0.02 for a person without the disease.

152. Use Scenario 5-7. If a randomly selected person is tested and the result is positive, the probability the individual has the disease is
A. 0.001.
B. 0.019.
C. 0.020.
D. 0.021.
E. 0.047.
Scenario 5-8
A student is chosen at random from the River City High School student body, and the following events are recorded:
M = The student is male
F = The student is female
B = The student ate breakfast that morning.
N = The student did not eat breakfast that morning.
The following tree diagram gives probabilities associated with these events.

153. Use Scenario 5-8. What is the probability that the student had breakfast?
A. 0.32
B. 0.40
C. 0.50
D. 0.64
E. 0.80

154. Use Scenario 5-8. Find \( P(B \mid F) \) and write in words what this expression represents.
A. 0.18; The probability the student ate breakfast and is female.
B. 0.18; The probability the student ate breakfast, given she is female.
C. 0.18; The probability the student is female, given she ate breakfast.
D. 0.30; The probability the student ate breakfast, given she is female.
E. 0.30; The probability the student is female, given she ate breakfast.

Scenario 5-9
You ask a sample of 370 people, “Should clinical trials on issues such as heart attacks that affect both sexes use subjects of just one sex?” The responses are in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>34</td>
<td>105</td>
</tr>
<tr>
<td>Female</td>
<td>46</td>
<td>185</td>
</tr>
</tbody>
</table>

Suppose you choose one of these people at random.

155. Use Scenario 5-9. What is the probability that the person is a woman, given that she said “Yes?”
A. 0.20
B. 0.22
C. 0.25
D. 0.50
E. 0.575

Scenario 5-10
The Venn diagram below describes the proportion of students who take chemistry and Spanish at Jefferson High School, Where A = Student takes chemistry and B = Students takes Spanish.

156. Use Scenario 5-10. Find the value of \( P(A \cup B) \) and describe it in words.
A. 0.1; The probability that the student takes both chemistry and Spanish.
B. 0.1; The probability that the student takes either chemistry or Spanish, but not both.
C. 0.5; The probability that the student takes either chemistry or Spanish, but not both.
D. 0.6; The probability that the student takes either chemistry or Spanish, or both.
E. 0.6; The probability that the student takes both chemistry and Spanish.
Scenario 5-11
The following table compares the hand dominance of 200 Canadian high-school students and what methods they prefer using to communicate with their friends.

<table>
<thead>
<tr>
<th></th>
<th>Cell phone/Text</th>
<th>In person</th>
<th>Online</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-handed</td>
<td>12</td>
<td>13</td>
<td>9</td>
<td>34</td>
</tr>
<tr>
<td>Right-handed</td>
<td>43</td>
<td>72</td>
<td>51</td>
<td>166</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>85</td>
<td>60</td>
<td>200</td>
</tr>
</tbody>
</table>

Suppose one student is chosen randomly from this group of 200.

157. Use Scenario 5-11. What is the probability that the student chosen is left-handed or prefers to communicate with friends in person?
A. 0.065
B. 0.17
C. 0.428
D. 0.53
E. 0.595

158. Use Scenario 5-11. Which of the following statements supports the conclusion that the event “Right-handed” and the event “Online” are not independent?
A. 0.51
B. 0.24
C. 0.34
D. 0.166
E. 0.095

Scenario 5-12
The letters p, q, r, and s represent probabilities for the four distinct regions in the Venn diagram below. For each question, indicate which expression describes the probability of the event indicated.

159. Use Scenario 5-12. $P(B | A)$
A. $s$
B. $s - r$
C. $q$
D. $r$
E. $\frac{r + s}{q + r + s}$

Scenario 5-13
One hundred high school students were asked if they had a dog, a cat, or both at home. Here are the results.

<table>
<thead>
<tr>
<th></th>
<th>Dog?</th>
<th>Cat?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>74</td>
<td>4</td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td>16</td>
</tr>
</tbody>
</table>

160. Use Scenario 5-13. If a single student is selected at random and you know she has a dog, what is the probability she also has a cat?
A. 0.04
B. 0.12
C. 0.22
D. 0.25
E. 0.75

161. Use Scenario 5-13. If two students are selected at random, what is the probability that neither of them has a dog or a cat?
A. 0.37
B. 0.540
C. 0.548
D. 0.655
E. 0.74

162. An ecologist studying starfish populations collects the following data on randomly-selected 1-meter by 1-meter plots on a rocky coastline:
- The number of starfish in the plot.
- The total weight of starfish in the plot.
- The percentage of area in the plot that is covered by barnacles (a popular food for starfish).
- Whether or not the plot is underwater midway between high and low tide.
How many of these measurements can be treated as continuous random variables and how many as discrete random variables?
A. Three continuous, one discrete.
B. Two continuous, two discrete.
C. One continuous, three discrete.
D. Two continuous, one discrete, and a fourth that cannot be treated as a random variable.
E. One continuous, two discrete, and a fourth that cannot be treated as a random variable.

163. A variable whose value is a numerical outcome of a random phenomenon is called
A. a random variable.
B. a parameter.
C. biased.
D. a random sample.
E. a statistic.

164. Which of the following is not a random variable?
A. The heights of randomly-selected buildings in New York City.
B. The suit of a card randomly-selected from a 52-card deck.
C. The number of children in randomly-selected households in the United States.
D. The amount of money won (or lost) by the next person to walk out of a casino in Las Vegas.
E. All of the above are random variables.

165. A random variable is
A. a hypothetical list of the possible outcomes of a random phenomenon.
B. any phenomenon in which outcomes are equally likely.
C. any number that changes in a predictable way in the long run.
D. a variable used to represent the outcome of a random phenomenon.
E. a variable whose value is a numerical outcome associated with a random phenomenon.
166. Suppose there are three balls in a box. On one of the balls is the number 1, on another is the number 2, and on the third is the number 3. You select two balls at random and without replacement from the box and note the two numbers observed. The sample space \( S \) consists of the three equally likely outcomes \{(1, 2), (1, 3), (2, 3)\}. Let \( X \) be the sum of the numbers on two balls selected. Which of the following is the correct probability distribution for \( X \)?

- A. \( \frac{1}{3} \) \( \frac{1}{3} \) \( \frac{1}{6} \) \( \frac{1}{3} \) \( \frac{1}{4} \)
- B. \( \frac{1}{3} \) \( \frac{1}{3} \) \( \frac{2}{6} \) \( \frac{2}{6} \) \( \frac{2}{6} \)
- C. \( \frac{1}{3} \) \( \frac{1}{3} \) \( \frac{2}{6} \) \( \frac{2}{6} \) \( \frac{1}{4} \)
- D. \( \frac{1}{3} \) \( \frac{1}{3} \) \( \frac{2}{6} \) \( \frac{2}{6} \) \( \frac{2}{6} \)
- E. \( \frac{1}{3} \) \( \frac{1}{3} \) \( \frac{2}{6} \) \( \frac{2}{6} \) \( \frac{1}{4} \)

Scenario 6-1

Flip a coin four times. If \( Z \) = the number of heads in four flips, then the probability distribution of \( Z \) is given in the table below.

<table>
<thead>
<tr>
<th>( Z )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(Z) )</td>
<td>0.0625</td>
<td>0.2500</td>
<td>0.3750</td>
<td>0.2500</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

167. Use Scenario 6-1. An expression the represents the probability of at least one tail is

- A. \( P(Z \geq 3) \)
- B. \( P(Z \leq 3) \)
- C. \( P(Z < 3) \)
- D. \( P(Z > 3) \)
- E. \( P(Z \geq 1) \)

Scenario 6-2

In a particular game, a fair die is tossed. If the number of spots showing is either 4 or 5 you win $1, if the number of spots showing is 6 you win $4, and if the number of spots showing is 1, 2, or 3 you win nothing. Let \( X \) be the amount that you win. Use Scenario 6-2. Which of the following is the expected value of \( X \)?

- A. $0.00
- B. $1.00
- C. $2.50
- D. $4.00
- E. $6.00

Scenario 6-3

In a population of students, the number of calculators a student owns is a random variable \( X \) described by the following probability distribution:

<table>
<thead>
<tr>
<th>( X )</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X) )</td>
<td>0.2</td>
<td>0.6</td>
<td>0.2</td>
</tr>
</tbody>
</table>

169. Use Scenario 6-3. Which of the following is the mean of \( X \)?

- A. 0.5
- B. 1
- C. 1.2
- D. 2
- E. The answer cannot be computed from the information given.

Scenario 6-4

<table>
<thead>
<tr>
<th>Number of cards</th>
<th>Payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$1,000</td>
</tr>
<tr>
<td>1000</td>
<td>$50</td>
</tr>
<tr>
<td>5000</td>
<td>$5</td>
</tr>
</tbody>
</table>

In the Florida scratch-card lottery, the numbers and values of prizes awarded for every 100,000 cards sold are

170. Use Scenario 6-4. The probability that a random scratch-card will pay off is

- A. .0250
- B. .0601
- C. .2500
- D. .6010
- E. .8500

Scenario 6-5

A small store keeps track of the number \( X \) of customers that make a purchase during the first hour that the store is open each day. Based on the records, \( X \) has the following probability distribution:

<table>
<thead>
<tr>
<th>( X )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X) )</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.6</td>
</tr>
</tbody>
</table>

171. Use Scenario 6-5. The mean number of customers that make a purchase during the first hour that the store is open is

- A. 2.0
- B. 2.5
- C. 2.9
- D. 3.0
- E. 4.0

172. Use Scenario 6-5. Consider the following game. You pay me an entry fee of \( x \) dollars; then I roll a fair die. If the die shows a number less than 3 I pay you nothing; if the die shows a 3 or 4, I give you back your entry fee of \( x \) dollars; if the die shows a 5, I will pay you $1; and if the die shows a 6, I pay you $3. What value of \( x \) makes the game fair (in terms of expected value) for both of us?

- A. $2
- B. $4
- C. $1
- D. $0.75
- E. $0.5

Scenario 6-6

The probability distribution of a continuous random variable \( X \) is given by the density curve below.

173. Use Scenario 6-6. The probability that \( X \) is between 0.5 and 1.5 is

- A. 1/4
- B. 1/3
- C. 1/2
- D. 3/4
- E. 1
Use Scenario 6-6. The probability that \( X = 1.5 \) is
A. 0.
B. very small; slightly larger than 0.
C. \( \frac{1}{4} \).
D. \( \frac{1}{3} \).
E. \( \frac{1}{2} \).

Use Scenario 6-9. The weights of grapefruits of a certain variety are approximately Normally distributed with a mean of 1 pound and a standard deviation of 0.12 pounds.

Use Scenario 6-7. Suppose \( X \) is a continuous random variable taking values between 0 and 2 and having the probability density function below.

\[
\begin{array}{c|c}
X & \text{Density} \\
0 & 0.5 \\
1 & 0.3 \\
2 & 0.2 \\
\end{array}
\]

P(\( X > 1.5 \)) has value
A. 0.50.
B. 0.33.
C. 0.25.
D. 0.125.
E. 0.0625.

Let the random variable \( X \) represent the profit made on a randomly selected day by a certain store. Assume \( X \) is Normal with a mean of $360 and standard deviation $50.

\[ P(X > 400) = \Phi\left(\frac{400 - 360}{50}\right) = \Phi(0.8) \]

A. 0.2881.
B. 0.8450.
C. 0.7881.
D. 0.2119.
E. 0.1600.

The weights of grapefruits of a certain variety are approximately Normally distributed with a mean of 1 pound and a standard deviation of 0.12 pounds.

Use Scenario 6-9. What is the probability that a randomly-selected grapefruit weighs more than 1.25 pounds?
A. 0.0188
B. 0.0156
C. 0.3156
D. 0.4013
E. 0.5987

Your friend Albert has invented a game involving two ten-sided dice. One of the dice has threes, fours, and fives on its faces, the other has sixes, eights, and tens. He won’t tell you how many of each number there are on the faces, but he does tell you that if \( X = \) rolls of the first die and \( Y = \) rolls of the second die, then

\( \mu_x = 3.6, \sigma_x = 0.5, \mu_y = 8.0, \sigma_y = 0.9 \). Let \( Z = \) the sum of the two dice when each is rolled once.

Use Scenario 6-10. What is the expected value of \( Z? \)
A. 1.7
B. 4.4
C. 8.8
D. 8.9
E. 11.6

Use Scenario 6-10. Here’s Albert’s game: You give him $10 each time you roll, and he pays you (in dollars) the amount that comes up on the dice. If \( P \) = the amount of money you gain each time you roll, and he pays you (in dollars) the amount that comes up on the dice, then

\( \mu_p = -1.6, \sigma_p = 1.45 \)

A. 1.6, \( \sigma_p = 1.45 \)
B. \( \mu_p = 1.6, \sigma_p = 1.45 \)
C. \( \mu_p = -1.6, \sigma_p = 1.2 \)
D. \( \mu_p = -1.6, \sigma_p = -1.38 \)
E. \( \mu_p = 1.6, \sigma_p = 13.8 \)

For which of the following counts would a binomial probability model be reasonable?
A. The number of traffic tickets written by each police officer in a large city during one month.
B. The number of hearts in a hand of five cards dealt from a standard deck of 52 cards that has been thoroughly shuffled.
C. The number of 7’s in a randomly selected set of five random digits from a table of random digits.
D. The number of phone calls received in a one-hour period.
E. All of the above.

Use Scenario 6-12. There are twenty multiple-choice questions on an exam, each having responses a, b, c, or d. Each question is worth five points and standard deviation $50.

\[ P(X > 400) = \Phi\left(\frac{400 - 360}{50}\right) = \Phi(0.8) \]

A. 0.2881.
B. 0.8450.
C. 0.7881.
D. 0.2119.
E. 0.1600.

An airplane has a front and a rear door that are both opened to allow passengers to exit when the plane lands. The plane has 100 passengers seated. The number of passengers exiting through the front door should have
A. a binomial distribution with mean 50.
B. a binomial distribution with 100 trials but success probability not equal to 0.5.
C. a geometric distribution with \( p = 0.5 \).
D. a normal distribution with a standard deviation of 5.
E. none of these.

In a certain game of chance, your chances of winning are 0.2. If you play the game five times and outcomes are independent, which of the following represents the probability that you win at least once?
A. \( 0.2^5 \)
B. \( 1 - (0.2)^5 \)
C. \( 1 - (0.2)^5 \)
D. \( \sum_{k=1}^{5} (0.2)^k (0.8)^{5-k} \)
E. \( 0.8^5 \)
186. Use Scenario 6-13. Which of the following is the approximate standard deviation \( \sigma \) of \( X \)?
   A. 0.40
   B. 0.24
   C. 19
   D. 360
   E. 9.20

187. Use Scenario 6-14. If you randomly select 20 bottles from those produced by this machine, what is the approximate probability that between 2 and 6 (inclusive) caps have been improperly applied?
   A. 0.19
   B. 0.26
   C. 0.38
   D. 0.74
   E. 0.92

188. Use Scenario 6-14. In a production run of 800 bottles, what is the standard deviation for the number of bottles with improperly applied caps?
   A. 1.38
   B. 6.16
   C. 6.32
   D. 6.89
   E. 8.72

189. A college basketball player makes \( \frac{5}{6} \) of his free throws. Assuming free throw attempts are independent, the probability that he makes exactly three of his next four free throws is:
   A. \( 4 \cdot \left( \frac{1}{6} \right)^3 \cdot \left( \frac{5}{6} \right)^2 \)
   B. \( \frac{1}{6} \cdot \left( \frac{5}{6} \right)^3 \)
   C. \( 3 \cdot \left( \frac{1}{6} \right)^2 \cdot \left( \frac{5}{6} \right)^3 \)
   D. \( \frac{1}{6} \cdot \left( \frac{5}{6} \right)^3 \)
   E. \( 4 \cdot \left( \frac{1}{6} \right)^2 \cdot \left( \frac{5}{6} \right)^3 \)

190. The binomial expression \( X \) gives the probability of
   A. at least 2 successes in 8 trials if the probability of success in one trial is 1/3.
   B. at least 2 successes in 8 trials if the probability of success in one trial is 2/3.
   C. exactly 2 successes in 8 trials if the probability of success in one trial is 1/3.
   D. exactly 2 successes in 8 trials if the probability of success in one trial is 2/3.
   E. at least 6 successes in 8 trials if the probability of success in one trial is 2/3.

191. A college basketball player makes 80% of her free throws. Suppose this probability is the same for each free throw she attempts, and free throw attempts are independent. The expected number of free throws required until she makes her first free throw of the season is:
   A. 2
   B. 1.25
   C. 0.80
   D. 0.31
   E. 0.13

192. Use Scenario 6-15. The expected value of \( X \) is:
   A. 1
   B. 1.5
   C. 2
   D. 2.5
   E. 3

193. Use Scenario 6-16. On average, how many people will the reporter have to stop before he finds his first Democrat?
   A. 1
   B. 1.33
   C. 1.67
   D. 2
   E. 2.33

194. Use Scenario 6-17. You’re going to give up and call a tow truck if you don’t find jumper cables by the time you’ve asked 10 people. What’s the probability you end up calling a tow truck?
   A. 0.8251
   B. 0.1749
   C. 0.1344
   D. 0.0333
   E. 0.0280

195. A jar has 250 marbles in it, 40 of which are red. What is the largest sample size we can take from the jar (without replacement) if we want to use the binomial distribution to model the number of red marbles in our sample?
   A. 50
   B. 40
   C. 25
   D. 4
   E. You can’t use a binomial distribution in this setting.

196. A survey conducted by Black Flag asked whether or not the action of a certain type of roach disk was effective in killing roaches. 79% of the respondents agreed that the roach disk was effective. The number 79% is a
   A. parameter.
   B. population.
   C. statistic.
   D. sample.
   E. sampling distribution.

Scenario 7-1
A CBS News/New York Times opinion poll asked 1,190 adults whether they would prefer balancing the Federal budget over cutting taxes; 59% of those asked answered “Yes.” Suppose that in fact 62% of all adults favor balancing the budget over cutting taxes.

197. Use Scenario 7-1. The number 62% is a
   A. evidence of bias.
   B. must be an error, since the actual percentage is 59%.
   C. a sampling distribution.
   D. is a statistic.
   E. is a parameter.

198. Which of the following is correct?
   A. parameters describe population characteristics
   B. parameters describe sample characteristics
   C. the population is a subset of the sample
   D. statistics must be based on a simple random sample
   E. both (A) and (D) are correct.
201. Use Scenario 7-2. The statistic that has the largest bias among these three is
A. statistic A. 
B. statistic B. 
C. statistic C. 
D. A and B have similar bias, and it is larger than the bias of C. 
E. B and C have similar bias, and it is larger than the bias of A.

202. Use Scenario 7-2. Based on the performance of the three statistics in many samples, which is preferred as an estimate of the parameter?
A. statistic A. 
B. statistic B. 
C. statistic C. 
D. either A or B would be equally good. 
E. either B or C would be equally good.

203. To reduce the variability of estimates from a simple random sample, you should 
A. use a smaller sample. 
B. increase the bias. 
C. use a count, not a percent. 
D. use a larger sample. 
E. use a percent, not a count.

204. “Congress passed a ban on the sale of assault weapons. Now there is a move to repeal that ban. Do you agree that the ban on sale of assault weapons should be repealed?” You ask that question to an SRS of 1000 adults in Texas (population 21 million people) and to a separate SRS of 1000 adults in Indiana (population 6 million people). The standard deviation of the sampling distribution for proportions in Indiana is approximately the same as in Texas, because the two SRSs are approximately the same size. 
A. larger than in Texas, because there are fewer people in Indiana. 
B. smaller than in Texas, because there are fewer people in Indiana. 
C. smaller than in Texas, because the sample is a larger proportion of the population. 
D. larger for \( n = n_1 \), and only the shape for the larger sample would be approximately Normal. 
E. Center and spread would be the same, but only the shape for the larger sample would be approximately Normal.

205. Use Scenario 7-3. The standard deviation \( \sigma_p \) of \( \hat{p} \) is approximately
A. 0.00121. 
B. 0.00187. 
C. 0.0110. 
D. 2.643. 
E. 2.683.

206. Use Scenario 7-3. If the true proportion is actually 0.03, the probability that \( \hat{p} \) is more than 0.054 is closest to
A. 0.000121. 
B. 0.0222. 
C. 0.0207. 
D. 0.0416. 
E. We can’t calculate this probability because we haven’t satisfied the Normality condition.

Scenario 7-4
According to a recent poll, 27% of Americans get 30 minutes of exercise at least five days each week. Let’s assume this is the parameter value for the population.

207. Use Scenario 7-4. If you take a simple random sample of 25 Americans and let \( \hat{p} \) = the proportion in the sample who get 30 minutes of exercise at least five days per week, what are the mean and standard deviation of the sampling distribution of \( \hat{p} \) ?
A. \( \mu_{\hat{p}} = 0.27; \sigma_{\hat{p}} = 0.0832 \) 
B. \( \mu_{\hat{p}} = 0.30; \sigma_{\hat{p}} = 0.0888 \) 
C. \( \mu_{\hat{p}} = 0.27; \sigma_{\hat{p}} = 0.0079 \) 
D. \( \mu_{\hat{p}} = 0.27; \sigma_{\hat{p}} = 0.0888 \) 
E. \( \mu_{\hat{p}} = 0.27; \sigma_{\hat{p}} = 0.1039 \)

208. Use Scenario 7-4. Suppose you increased the sample size in the previous question to \( n = 50 \). How would the sampling distribution of \( \hat{p} \) compare to the sampling distribution for \( n = 25 \)?
A. Center and spread would be the same, both distributions would be approximately Normal. 
B. Center and spread would be the same, but only the shape for the larger sample would be approximately Normal. 
C. Center would be the same, spread would be smaller for \( n = 50 \), both distributions would be approximately Normal. 
D. Center would be the same, spread would be smaller for \( n = 50 \), and only the shape for the larger sample would be approximately Normal. 
E. Center would be the same, spread would be larger for \( n = 50 \), and only the shape for the larger sample would be approximately Normal.

209. A fair coin (one for which both the probability of heads and the probability of tails are 0.5) is tossed 60 times. The probability that less than 1/3 of the tosses are heads is closest to
A. 0.0001. 
B. 0.0222. 
C. 0.0110. 
D. 0.09. 
E. 0.0049.

210. A college basketball player makes 80% of his free throws. Over the course of the season she will attempt 100 free throws. Assuming free throw attempts are independent, the probability that the number of free throws she makes exceeds 80 is approximately
A. 0.2000. 
B. 0.2266. 
C. 0.5000. 
D. 0.7734. 
E. 0.8000.