Chapter 1: Exploring Data

___ 1. As part of survey of college students a researcher is interested in the variable class standing. She records a 1 if the student is a freshman, a 2 if the student is a sophomore, a 3 if the student is a junior, and a 4 if the student is a senior. The variable class standing is
   a) categorical.  b) numerical.  c) quantitatively categorical.  d) all of the above.

___ 2. A survey records many variables of interest to the researchers conducting the survey. Which of the following variables, from a survey conducted by the U.S. Postal Service, 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.

___ 3. 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. Which of the following 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.”

___ 4. A description of different houses on the market includes the following three variables. Which of these variables is quantitative?
   a) the square footage of the house.  b) the monthly gas bill.  c) the monthly electric bill.  d) all of the above.

___ 5. A professor records the values of several variables for each student in her class. Which of the following 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.

___ 6. 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 the bars are true rectangles.
   c) label the vertical axis, so the reader can determine the counts or percent in each class interval.
   d) make certain the mean and median are contained in the same class interval, so the correct type of skewness can be identified.

___ 7. In drawing a histogram, which of the following suggestions should be followed?
   a) Leave large gaps between bars. This allows room for comments.
   b) The heights of bars should equal the class frequency.
   c) Generally, bars should be square so that both the height and width equal the class count.
   d) The scale of the vertical axis should be that of the variable whose distribution you are displaying.
Chapter 1 Review

Each of the following two histograms represents the distribution of acceptance rates (percent accepted) among 25 business schools in 1995. The histograms use different class intervals but are based on the same data. In each class interval, the left endpoint is included but not the right.

8. What percent of the schools have an acceptance rate of less than 20%?
   a) 3%.  b) 4%.  c) 12%.  d) 16%.

9. Which interval contains fewer than half of all the observations?
   a) 20% = acceptance rate < 35%.
   b) 22.5% = acceptance rate < 37.5%.
   c) 25% = acceptance rate < 40%.
   d) 30% = acceptance rate < 45%.
In a statistics class with 136 students, the professor records how much money each student has in his or her possession during the first class of the semester. The histogram below is of the data collected.

---

10. The percentage of students with under $10 in their possession is closest to
   a) 35%.   b) 50%.   c) 60%.   d) 70%.

11. The histogram
   a) is skewed right.   b) has an outlier.   c) is asymmetric.   d) all of the above.

12. The number of students with over $30 in their possession is
   a) less than 5.   b) about 10.   c) about 30.   d) more than 100.

Below is a histogram of the gold medal-winning high jumps for the Olympic Games.

---

13. The mean of this histogram is approximately
   a) 75 inches.   b) 77.5 inches.   c) 82 inches.   d) 90 inches.

14. The percentage of these winning jumps that were at least 7 feet (84 inches) is about
   a) 9%.   b) 14%.   c) 23%.   d) 37%.
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.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

15. This stemplot is most similar to
   a) a histogram with class intervals 110 < score < 120, 120 < score < 130, etc.
   b) a timeplot of the data with the observations taken in increasing order.
   c) a boxplot of the data.
   d) reporting the five-point summary for the data, with the mean.

16. To which of the following data sets does the stemplot correspond?
   a) all integers between 116 and 179.
   b) 1, 2, 3, 4, 6, 7, 8, 9.
   c) 16, 18, 21, 24, 28, 33, 37, 42, 46, 79.
   d) 116, 118, 121, 124, 128, 133, 137, 142, 146, 179.

17. The median point total for this class is
   a) 130.
   b) 130.5.
   c) 133.
   d) 137.

During the early part of the 1994 baseball season, many sports fans and baseball players noticed that the number of home runs being hit seemed to be unusually large. Below are the team-by-team statistics on home runs hit through Friday, June 3, 1994 (from the *Columbus Dispatch* Sports Section, Sunday, June 5, 1994), in the form of separate stemplots for the number of home runs by American and National League teams.

<table>
<thead>
<tr>
<th>American League</th>
<th>National League</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0 3 9</td>
</tr>
<tr>
<td>5</td>
<td>1 4 7 8 8</td>
</tr>
<tr>
<td>6</td>
<td>4 8 8</td>
</tr>
<tr>
<td>7</td>
<td>5 7</td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

18. The median for the number of home runs for the American League teams is
   a) lower than that for the National League teams.
   b) 45.
   c) 50.
   d) 57.5.

19. Which of the following is a correct statement?
   a) The American League plot is reasonably symmetric.
   b) The National League plot is slightly skewed to the left.
   c) The median number of home runs hit by American League teams was higher than that of National League teams.
   d) All of the above.
20. Which of the following four bar graphs is equivalent to the pie chart in the reference?

a) 

b) 

c) 

d)
21. Below is a bar graph of class standing for a seminar containing seven students who are either freshman, sophomores, juniors, or seniors. (The bar for the juniors has been omitted.)

![Bar graph of class standing](image)

The number of students in the seminar who are juniors
a) is 1.  b) is 2.  c) is 3.  d) cannot be determined from the information given.

22. The timeplot below gives the number of burglaries committed each month for a city in Ohio. The plot is for the three-year period January 1987–December 1989.

![Timeplot of burglaries](image)

The number of burglaries in December 1989, the last date recorded in the timeplot, was about
a) 22.  b) 27.  c) 32.  d) 37.

23. Which of the following is a true statement?
   a) The number of burglaries in each month of 1988 was lower than the number of burglaries in each month of 1989.
   b) The median number of burglaries for a month in 1988 was a little over 25.
   c) The total number of burglaries in 1989 was higher than in 1988.
   d) None of the above.

24. The maximum number of burglaries for a month in 1988 was about
   a) 20.  b) 25.  c) 30.  d) 35.

25. 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) $10.  b) $15.  c) $20.  d) $25.
26. The average salary of all female workers is $35,000. The average salary of all male workers is $41,000. What must be true about the average salary of all workers?
   a) It must be $38,000.
   b) It must be larger than the median salary.
   c) It could be any number between $35,000 and $41,000.
   d) It must be larger than $38,000.

27. 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.

   28  31  34  35  37  41  42  42  42  47
   49  51  52  52  60  61  67  72  75  77

Suppose each employee in the company receives a $3000 raise for next year (each employee's salary is increased by $3000). The mean salary for the employees will
   a) be unchanged.
   b) increase by $3000.
   c) be multiplied by $3000.
   d) increase by \( \sqrt{3000} \).

28. The ages of people in a class (to the nearest year) are as follows:

   Age 18 19 20 21 22 23 24 25 32
   Number of Students 14 120 200 200 90 30 10 2 1

   What is true about the median age?
   a) It must be 20.
   b) It could be any number between 19 and 21.
   c) It must be 21.
   d) It must be over 21.

29. A sample of serum cholesterol levels of six men who visited a cholesterol screening clinic located in the downtown area yielded values of

   218  273  210  259  290  232

The mean cholesterol level in this sample is
   a) 229.5.
   b) 245.5.
   c) 247.0.
   d) 266.0.

30. A sample of five recent births at a local hospital yielded birthweights (in ounces) of

   89  122  137  144  98

The mean birthweight in this sample is
   a) 118.
   b) 122.
   c) 137.
   d) 141.

31. The mean age of five people in a room is 30 years. One of the people, whose age is 50 years, leaves the room. The mean age of the remaining four people in the room
   a) is 40.
   b) is 30.
   c) is 25.
   d) cannot be determined from the information given.

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.

   28  31  34  35  37  41  42  42  42  47
   49  51  52  52  60  61  67  72  75  77

32. The median salary of the 20 employees is
   a) $49,000.
   b) $48,000.
   c) $47,000.
   d) $42,000.

33. Suppose each employee in the company receives a $3000 raise for next year (each employee's salary is increased by $3000). The median salary for the employees working for the company will
   a) be unchanged.
   b) increase by $3000.
   c) be multiplied by $3000.
   d) increase by \( \sqrt{3000} \).
34. A sample of five recent births at a local hospital yielded birthweights (in ounces) of
89 122 137 144 98
The median birthweight in this sample is
a) 118. b) 122. c) 137. d) 141.

35. A sample of serum cholesterol levels of six men who visited a cholesterol screening clinic located in the
downtown area yielded values of
218 273 210 259 290 232
The median cholesterol level in this sample is
a) 229.5. b) 245.5. c) 247.0. d) 266.0.

36. The median age of five people in a room is 30 years. One of the people, whose age is 50 years, leaves
the room. The median age of the remaining four people in the room
a) is 40. b) is 30. c) is 25. d) cannot be determined from the information given.

37. A set of data has a median that is much larger than the mean. Which of the following statements is
most consistent with this information?
a) A stemplot of the data is symmetric.
b) A stemplot of the data is skewed left.
c) A stemplot of the data is skewed right.
d) The data set must be so large that it would be better to draw a histogram than a stemplot.

38. In a class of 100 students, the grades on a statistics test are summarized in the following frequency
table.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>91-100</td>
<td>11</td>
</tr>
<tr>
<td>81-90</td>
<td>31</td>
</tr>
<tr>
<td>71-80</td>
<td>42</td>
</tr>
<tr>
<td>61-70</td>
<td>16</td>
</tr>
</tbody>
</table>

The median grade is in which of the following intervals?
a) 61-70. b) 71-80. c) 81-90. d) 91-100.

39. A reporter wishes to portray baseball players as overpaid. Which measure of center should he report as
the average salary of major league players?
a) the mean.
b) the median.
c) either the mean or median; they will be equal in this case.
d) neither the mean nor the median; both will be much lower than the actual average salary.

40. The number of new projects started each month at an advertising agency for the last six months is
2 5 3 3 6 3
The interquartile range for the above data is
a) 1.0. b) 4.0. c) 5.0. d) 2.0.

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.

28 31 34 35 37 41 42 42 42 47
49 51 52 52 60 61 67 72 75 77

41. The first quartile of the 20 salaries is
a) $35,000. b) $36,000. c) $37,000. d) $39,000.

42. The interquartile range of the 20 salaries is
a) $19,000. b) $19,500. c) $21,500. d) $49,000.
43. Suppose each employee in the company receives a $3000 raise for next year (each employee's salary is increased by $3000). The interquartile range of the salaries for the employees will
a) be unchanged.  
 b) increase by $3000.  
 c) be multiplied by $3000.  
 d) increase by $\sqrt{3000}$.

44. The Insurance Institute for Highway Safety publishes data on the total damage suffered by compact automobiles in a series of controlled, low-speed collisions. A sample of the data in dollars, with brand names removed, is
1000  600  800  1000

The interquartile range of the data is
a) 300.  
 b) 200.  
 c) 400.  
 d) none of the above.

45. In a statistics class with 136 students, the professor records how much money each student has in his or her possession during the first class of the semester. The histogram below is of the data collected.

From the histogram, which of the following is true?
 a) The mean is much larger than the median.  
 b) The mean is much smaller than the median.  
 c) The mean and median are approximately equal.  
 d) It is impossible to compare the mean and median for these data.

46. The five-number summary of a set of data is
 a) the minimum, first quartile, median, third quartile, and maximum.  
 b) the mean, median, mode, variance, and standard deviation.  
 c) any five-digit number that describes the data.  
 d) any five single-digit numbers that are measures of center and spread.
A sample was taken of the salaries of 20 employees of a large company. The following is a boxplot of the salaries (in thousands of dollars) for this year.

47. Based on this 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 interquartile range is about $20,000.
   d) The median salary is about $40,000.

48. Based on this boxplot, which of the following statements is true?
   a) The salary distribution is fairly symmetric.
   b) About 10 employees make more than $50,000.
   c) Nobody makes more than $80,000.
   d) All of the above.

49. Based on this 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.

The boxplot below is of the birthweights (in ounces) of a sample of 160 infants born in a local hospital.

50. The median birthweight is approximately
    a) 90 ounces.  b) 100 ounces.  c) 110 ounces.  d) 120 ounces.

51. About 40 of the birthweights were less than
    a) 92 ounces.  b) 102 ounces.  c) 112 ounces.  d) 122 ounces.

52. The number of children with birthweights between 100 and 120 ounces is approximately
    a) 40.  b) 50.  c) 80.  d) 100.
53. This is a standard deviation contest. Which of the following sets of four numbers has the largest possible standard deviation?
   a) 7, 8, 9, 10.   b) 5, 5, 5, 5.   c) 0, 0, 10, 10.   d) 0, 1, 2, 3.

54. 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.

   28 31 34 35 37 41 42 42 42 47
   49 51 52 52 60 61 67 72 75 77

Suppose each employee in the company receives a $3000 raise for next year (each employee's salary is increased by $3000). The standard deviation of the salaries for the employees will
   a) be unchanged.   c) be multiplied by $3000.
   b) increase by $3000.   d) increase by $\sqrt{3000}$.

55. The standard deviation of 16 measurements of people's weights (in pounds) is computed to be 5.4. The variance of these measurements is
   a) 2.24.   b) 9.16.   c) 52.34.   d) 256.

56. The standard deviation of 16 measurements of people's 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.

57. The Insurance Institute for Highway Safety publishes data on the total damage suffered by compact automobiles in a series of controlled, low-speed collisions. A sample of the data in dollars, with brand names removed, is

   1000 600 800 1000
   1000 800 600 1000

The variance of the above data is
   a) 191.5.   b) 36,667.   c) 165.8.   d) 27,500.

58. The rental values (in dollars) of a sample of four available apartments close to the university are

   470 600 580 550

The standard deviation of the sample is
   a) $30.31.   b) $35.   c) $57.15.   d) $49.50.

Answer Key

1. a 16. d 31. c 46. a
2. a 17. b 32. b 47. c
3. c 18. d 33. b 48. d
4. d 19. d 34. b 49. a
5. b 20. b 35. b 50. c
6. c 21. c 36. d 51. b
7. b 22. a 37. b 52. c
8. d 23. c 38. b 53. c
9. d 24. d 39. a 54. a
10. b 25. a 40. d 55. b
11. d 26. c 41. d 56. c
12. b 27. b 42. c 57. b
13. c 28. a 43. a 58. c
14. d 29. c 44. a
15. a 30. a 45. a