

Ch. 2 Methods for Describing Sets of Data

2.1 Describing Qualitative Data

1 Identify Classes/Compute Class Frequencies/Relative Frequencies/Percentages

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) In an eye color study, 25 out of 50 people in the sample had brown eyes. In this situation, what does the number .50 represent?
 - A) a class relative frequency
 - B) a class
 - C) a class frequency
 - D) a class percentage

- 2) What class percentage corresponds to a class relative frequency of .37?
 - A) 37%
 - B) .37%
 - C) .63%
 - D) 63%

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 3) A sample of 100 e-mail users were asked whether their primary e-mail account was a free account, an institutional (school or work) account, or an account that they pay for personally. Identify the classes for the resulting data.

2 Construct Frequency/Relative Frequency Table

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 4) What number is missing from the table?

Grades on Test	Frequency	Relative Frequency
A	6	.24
B	7	
C	9	.36
D	2	.08
F	1	.04

- A) .28
- B) .07
- C) .72
- D) .70

- 5) What number is missing from the table?

Year in College	Frequency	Relative Frequency
Freshman	600	.30
Sophomore	560	.28
Junior		.22
Senior	400	.20

- A) 440
- B) 220
- C) 480
- D) 520

9)



The manager of a store conducted a customer survey to determine why customers shopped at the store. The results are shown in the figure. What proportion of customers responded that merchandise was the reason they shopped at the store?

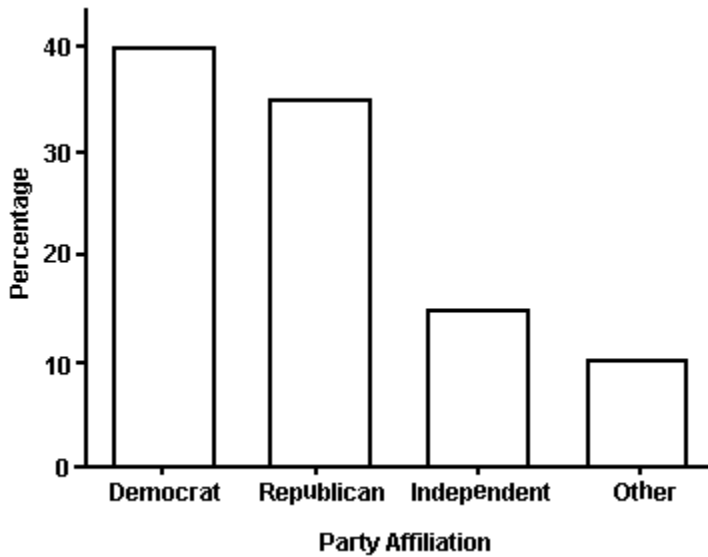
A) $\frac{3}{7}$

B) 30

C) $\frac{1}{2}$

D) $\frac{2}{7}$

10)



The bar graph shows the political affiliation of 1000 registered U.S. voters. What percentage of the voters belonged to one of the traditional two parties (Democratic or Republican)?

A) 75%

B) 40%

C) 35%

D) 25%

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 11) The data below show the types of medals won by athletes representing the United States in the Winter Olympics.

gold	gold	silver	gold	bronze	silver	silver
bronze	gold	silver	silver	bronze	silver	gold
gold	silver	silver	bronze	bronze	gold	silver
gold	gold	bronze	bronze			

- Construct a frequency table for the data.
- Construct a relative frequency table for the data.
- Construct a frequency bar graph for the data.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

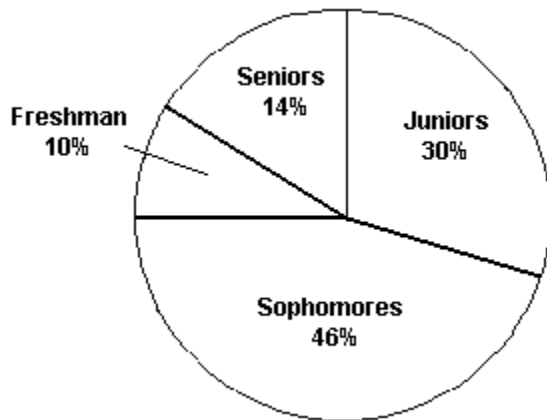
- 12) The bars in a bar graph can be arranged by height in ascending order from left to right.
A) True B) False
- 13) Either vertical or horizontal bars can be used when constructing a bar graph.
A) True B) False

4 Construct, Interpret Pie Chart

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

14)

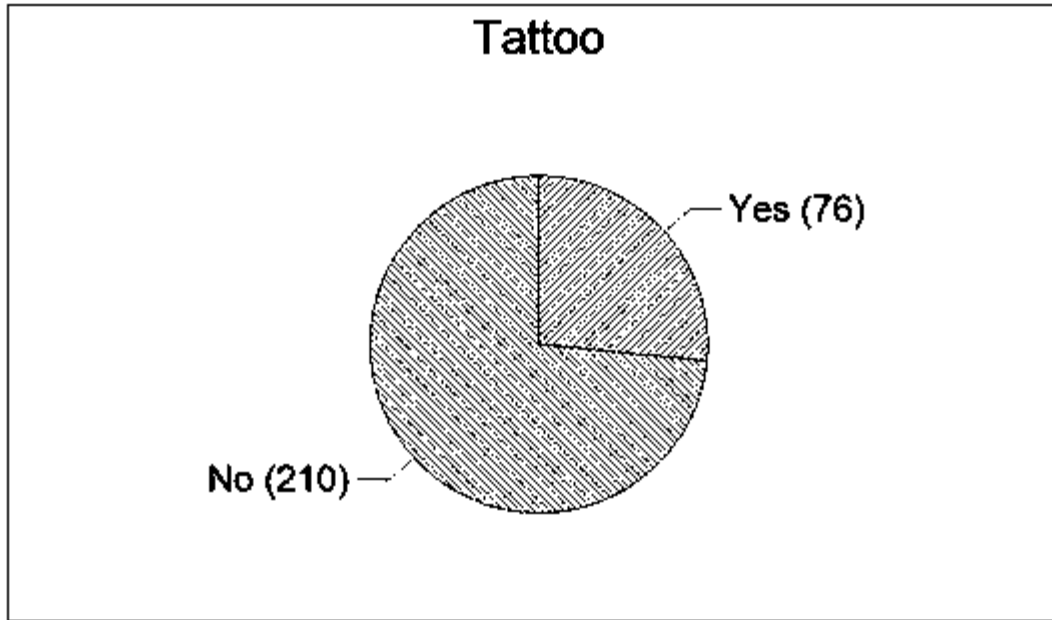


The pie chart shows the classifications of students in a statistics class.

What percentage of the class consists of freshman, sophomores, and juniors?

- A) 86% B) 14% C) 44% D) 54%

- 15) One of the questions posed to a sample of 286 incoming freshmen at a large public university was, "Do you have any tattoos?" Their responses are shown below in the pie chart. Please note that the values shown represent the number of responses in each category.



Based on the responses shown in the pie chart, what percentage of the freshmen responded with "Yes?"

A) 76 B) 76% C) 26.6% D) 73.4%

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 16) The table shows the number of each type of book found at an online auction site during a recent search.

Type of Book	Number
Children's	51,033
Fiction	141,114
Nonfiction	253,074
Educational	67,252

- Construct a relative frequency table for the book data.
- Construct a pie chart for the book data.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- 17) If 25% of your statistics class is sophomores, then in a pie chart representing classifications of the students in your statistics class the slice assigned to sophomores is 90° .
- A) True B) False
- 18) The slices of a pie chart must be arranged from largest to smallest in a clockwise direction.
- A) True B) False

5 Construct, Interpret Pareto Diagram

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

- 19) What characteristic of a Pareto diagram distinguishes it from other bar graphs?

20) The table shows the number of each type of car sold in June.

Car	Number
compact	7,204
sedan	9,089
small SUV	20,418
large SUV	13,691
minivan	15,837
truck	15,350
Total	81,589

- Construct a relative frequency table for the car sales.
- Construct a Pareto diagram for the car sales using the class percentages as the heights of the bars.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- Class relative frequencies must be used, rather than class frequencies or class percentages, when constructing a Pareto diagram.
 - True
 - False
- A Pareto diagram is a pie chart where the slices are arranged from largest to smallest in a counterclockwise direction.
 - True
 - False

6 Construct, Interpret Side-by-Side Bar Chart

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

- An annual survey sent to retail store managers contained the question "Did your store suffer any losses due to employee theft?" The responses are summarized in the table for two years. Compare the responses for the two years using side-by-side bar charts. What inferences can be made from the charts?

Employee Theft	Percentage in year 1	Percentage in year 2
Yes	34	23
No	51	68
Don't know	15	9
Totals	100	100

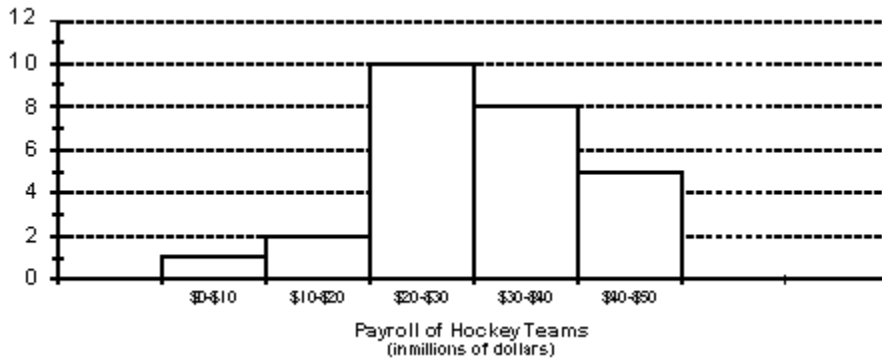
2.2 Graphical Methods for Describing Quantitative Data

1 Construct, Interpret Histogram

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) The payroll amounts for all teams in an international hockey league are shown below using a graphical technique from chapter 2 of the text. How many of the hockey team payrolls exceeded \$20 million (Note: Assume that no payroll was exactly \$20 million)?



- A) 8 teams B) 23 teams C) 10 teams D) 18 teams

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 2) The data show the total number of medals (gold, silver, and bronze) won by each country winning at least one gold medal in the Winter Olympics.

1 2 3 3 4 9 9 11 11

11 14 14 19 22 23 24 25 29

- a. Complete the class frequency table for the data.

Total Medals	Frequency
1-5	
6-10	
11-15	
16-20	
21-25	
26-30	

- b. Using the classes from the frequency table, construct a histogram for the data.

- 3) The total points scored by a basketball team for each game during its last season have been summarized in the table below.

Score	Frequency
41–60	3
61–80	8
81–100	12
101–120	7

- Explain why you cannot use the information in the table to construct a stem-and-leaf display for the data.
- Construct a histogram for the scores.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- All class intervals in a histogram have the same width.
 - True
 - False
- A histogram can be constructed using either class frequencies or class relative frequencies as the heights of the bars.
 - True
 - False
- The bars in a histogram should be arranged by height in descending order from left to right.
 - True
 - False

2 Construct, Interpret Stem-and-Leaf Display

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- A survey was conducted to determine how people feel about the quality of programming available on television. Respondents were asked to rate the overall quality from 0 (no quality at all) to 100 (extremely good quality). The stem-and-leaf display of the data is shown below.

Stem	Leaf
3	3 6
4	0 3 4 7 8 9 9 9
5	0 1 1 2 3 4 5
6	1 2 5 6 6
7	5 8
8	
9	3

What percentage of the respondents rated overall television quality as very good (regarded as ratings of 80 and above)?

- 4%
- 1%
- 12%
- 3%

8) 252 randomly sampled college students were asked, among other things, to estimate their college grade point average (GPA). The responses are shown in the stem-and-leaf plot shown below. Notice that a GPA of 3.65 would be indicated with a stem of 36 and a leaf of 5 in the plot. How many of the students who responded had GPA's that exceeded 3.55?

Stem and Leaf Plot of GPA

Leaf Digit Unit = 0.01
 19 9 represents 1.99

Minimum 1.9900
 Median 3.1050
 Maximum 4.0000

	Stem	Leaves
1	19	9
5	20	0668
6	21	0
11	22	05567
15	23	0113
20	24	00005
33	25	00000000000067
46	26	0000005577789
61	27	000000134455578
79	28	000000000144667799
88	29	002356777
116	30	000000000000000000011344559
(19)	31	0000000000112235666
117	32	0000000000000000345568
95	33	00000000025557
80	34	000000000000000333444566677889
49	35	000003355566677899
31	36	000005
25	37	022235588899
13	38	00002579
5	39	7
4	40	0000

252 cases included

- A) 31 B) 49 C) 39 D) 19

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

9) The scores for a statistics test are as follows:

87 76 91 77 94 96 88 85 66 89
 79 96 54 98 83 88 82 56 18 69

Create a stem-and-leaf display for the data.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

10) For large data sets, a stem-and-leaf display is a better choice than a histogram.

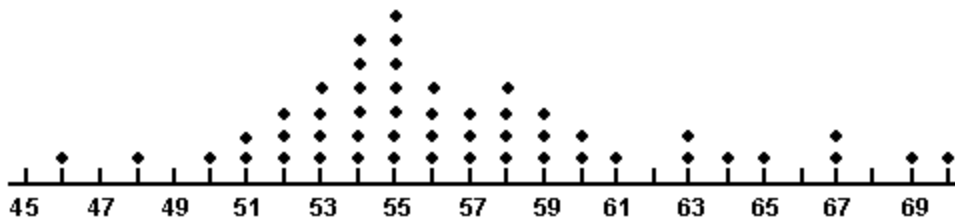
- A) True B) False

3 Construct, Interpret Dot-Plot

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

11) A dot plot of the speeds of a sample of 50 cars passing a policeman with a radar gun is shown below.



What proportion of the motorists were driving above the posted speed limit of 60 miles per hour?

- A) 0.18 B) 0.22 C) 0.04 D) 2
- 12) Which of the graphical techniques below can be used to summarize qualitative data?
A) bar graph B) dot plot
C) stem-and-leaf plot D) box plot
- 13) Parking at a university has become a problem. University administrators are interested in determining the average time it takes a student to find a parking spot. An administrator inconspicuously followed 300 students and recorded how long it took each of them to find a parking spot. Which of the following types of graphs should not be used to display information concerning the students parking times?
A) pie chart B) stem-and-leaf display
C) histogram D) box plot
- 14) Fill in the blank. One advantage of the _____ is that the actual data values are retained in the graphical summarization of the data.
A) stem-and-leaf plot B) histogram C) pie chart

2.3 Summation Notation

1 Use Summation Notation

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) A data set contains the observations 2, 6, 4, 7, 3. Find $\sum x$.
A) 22 B) 19 C) 1008 D) 5
- 2) A data set contains the observations 8, 1, 6, 7, 2. Find $(\sum x)^2$.
A) 576 B) 154 C) 48 D) 24
- 3) A data set contains the observations 5, 3, 6, 1, 4. Find $\sum x^2$.
A) 87 B) 361 C) 38 D) 19
- 4) A data set contains the observations 8, 3, 2, 6, 7. Find $\sum(x - 5)$.
A) 1 B) 51 C) 31 D) 21

- 5) A data set contains the observations 1, 6, 4, 5, 7. Find $\sum x^2 - \frac{(\sum x)^2}{5}$.
- A) 21.2 B) 423.2 C) 101.6 D) 232.8
- 6) Which expression represents the sum of the squares of the observations in a data set?
- A) $\sum x^2$ B) $(\sum x^2)^2$ C) $\sum \sqrt{x}$ D) $\sqrt{\sum x}$

2.4 Numerical Measures of Central Tendency

1 Find Mean, Median, Mode

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) The amount spent on textbooks for the fall term was recorded for a sample of five university students - \$400, \$350, \$600, \$525, and \$450. Calculate the value of the sample mean for the data.
- A) \$450 B) \$465 C) \$600 D) \$400
- 2) The amount spent on textbooks for the fall term was recorded for a sample of five university students - \$400, \$350, \$600, \$525, and \$450. Calculate the value of the sample median for the data.
- A) \$450 B) \$465 C) \$600 D) \$400
- 3) A sociologist recently conducted a survey of senior citizens who have net worths too high to qualify for Medicaid but have no private health insurance. The ages of the 25 uninsured senior citizens were as follows:

72	77	70	80	90
78	65	93	69	94
73	96	80	66	85
67	72	85	74	77
64	91	79	68	86

Find the median of the observations.

- A) 77 B) 74 C) 78 D) 77.5
- 4) The scores for a statistics test are as follows:

95	76	68	77	73	92	81	85	77	89
79	87	50	60	85	72	85	62	18	74

Compute the mean score.

- A) 74.25 B) 77.45 C) 65.20 D) 75

5) A shoe retailer keeps track of all types of information about sales of newly released shoe styles. One newly released style was marketed to tall people. Listed below are the shoe sizes of 12 randomly selected customers who purchased the new style. Find the mode of the shoe sizes.

- | | | | |
|----------------|-----------------|----------------|-----------------|
| $9\frac{1}{2}$ | 11 | 12 | $11\frac{1}{2}$ |
| $8\frac{1}{2}$ | $10\frac{1}{2}$ | 8 | 11 |
| 10 | 11 | $9\frac{1}{2}$ | 10 |

- A) 11 B) $10\frac{1}{4}$ C) $10\frac{1}{2}$ D) $9\frac{1}{2}$

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

6) Each year advertisers spend billions of dollars purchasing commercial time on network television. In the first 6 months of one year, advertisers spent \$1.1 billion. Who were the largest spenders? In a recent article, the top 10 leading spenders and how much each spent (in million of dollars) were listed:

Company A	\$70.4	Company F	\$24.6
Company B	61.7	Company G	24.3
Company C	57.2	Company H	23.1
Company D	54.5	Company I	21.8
Company E	31.4	Company J	19.9

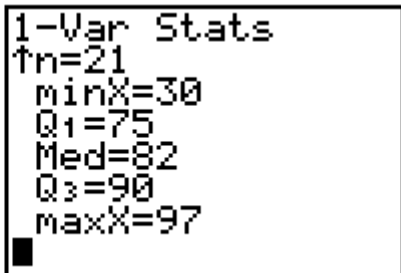
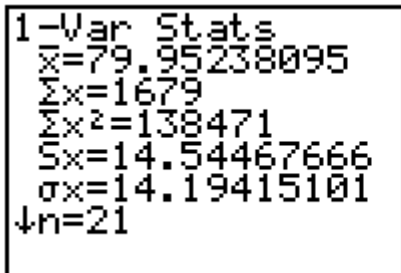
Calculate the mean and median for the data.

7) The data show the total number of medals (gold, silver, and bronze) won by each country winning at least one gold medal in the Winter Olympics. Find the mean, median, and mode of the numbers of medals won by these countries.

- | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 3 | 3 | 4 | 9 | 9 | 11 | 11 |
| 11 | 14 | 14 | 19 | 22 | 23 | 24 | 25 | 29 |

8) Calculate the mean of a sample for which $\sum x = 196$ and $n = 8$.

9) The calculator screens summarize a data set.



- How many data items are in the set?
- What is the sum of the data?
- Identify the mean, median, and mode, if possible.

2 Interpret Measures of Central Tendency

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

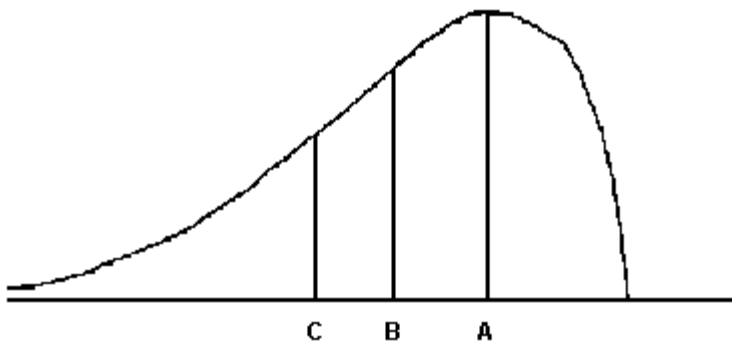
- 10) At the U.S. Open Tennis Championship a statistician keeps track of every serve that a player hits during the tournament. The statistician reported that the mean serve speed of a particular player was 95 miles per hour. Suppose that the statistician indicated that the serve speed distribution was skewed to the left. Which of the following values is most likely the value of the median serve speed?
A) 102 mph B) 88 mph C) 81 mph D) 95 mph
- 11) The amount spent on textbooks for the fall term was recorded for a sample of five hundred university students. The mean expenditure was calculated to be \$500 and the median expenditure was calculated to be \$425. Which of the following interpretations of the mean is correct?
A) The average of the textbook costs sampled was \$500
B) The most frequently occurring textbook cost in the sample was \$500
C) 50% of the students sampled had textbook costs equal to \$500
D) 50% of the students sampled had textbook costs that were less than \$500
- 12) The amount spent on textbooks for the fall term was recorded for a sample of five hundred university students. The mean expenditure was calculated to be \$500 and the median expenditure was calculated to be \$425. Which of the following interpretations of the median is correct?
A) The average of the textbook costs sampled was \$425
B) The most frequently occurring textbook cost in the sample was \$425
C) 50% of the students sampled had textbook costs equal to \$425
D) 50% of the students sampled had textbook costs that were less than \$425
- 13) During one recent year, U.S. consumers redeemed 6.79 billion manufacturers' coupons and saved themselves \$2.56 billion. Calculate and interpret the mean savings per coupon.
A) The average savings was \$0.38 per coupon.
B) The average savings was 265.2 cents per coupon.
C) Half of all coupons were worth more than 265.2 cents in savings.
D) Half of all coupons were worth more than \$0.38 in savings.
- 14) The output below displays the mean and median for the state high school dropout rates in year 1 and in year 5.

	Year 1	Year 5
N	51	51
MEAN	28.26	26.26
MEDIAN	27.41	25.28

Interpret the year 5 median dropout rate of 25.28.

- A) Half of the 51 states had a dropout rate below 25.28%.
B) Most of the 51 states had a dropout rate close to 25.28%.
C) The most frequently observed dropout rate of the 51 states was 25.28%.
D) Half of the 51 states had a dropout rate of 25.28%.

15)



For the distribution drawn here, identify the mean, median, and mode.

- A) A = mode, B = median, C = mean
B) A = median, B = mode, C = mean
C) A = mode, B = mean, C = median
D) A = mean, B = mode, C = median
- 16) In a distribution that is skewed to the right, what is the relationship of the mean, median, and mode?
A) mean > median > mode
B) median > mean > mode
C) mode > median > mode
D) mode > mean > median
- 17) Many firms use on-the-job training to teach their employees computer programming. Suppose you work in the personnel department of a firm that just finished training a group of its employees to program, and you have been requested to review the performance of one of the trainees on the final test that was given to all trainees. The mean of the test scores is 74. Additional information indicated that the median of the test scores was 84. What type of distribution most likely describes the shape of the test scores?
A) skewed to the left
B) symmetric
C) skewed to the right
D) unable to determine with the information given
- 18) A shoe company reports the mode for the shoe sizes of men's shoes is 12. Interpret this result.
A) The most frequently occurring shoe size for men is size 12
B) Most men have shoe sizes between 11 and 13.
C) Half of the shoes sold to men are larger than a size 12
D) Half of all men's shoe sizes are size 12
- 19) Which of the following is *not* a measure of central tendency?
A) range
B) median
C) mode
D) mean
- 20) The distribution of salaries of professional basketball players is skewed to the right. Which measure of central tendency would be the best measure to determine the location of the center of the distribution?
A) median
B) mode
C) mean
D) range

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 21) Parking at a university has become a problem. University administrators are interested in determining the average time it takes a student to find a parking spot. An administrator inconspicuously followed 280 students and recorded how long it took each of them to find a parking spot. The times had a distribution that was skewed to the left. Based on this information, discuss the relationship between the mean and the median for the 280 times collected.

22) The output below displays the mean and median for the state high school dropout rates in year 1 and in year 5.

	Year 1	Year 5
N	51	51
MEAN	28.55	26.19
MEDIAN	27.85	25.82

Use the information to determine the shape of the distributions of the high school dropout rates in year 1 and year 5.

23) The total points scored by a basketball team for each game during its last season have been summarized in the table below. Identify the modal class of the distribution of scores.

Score	Frequency
41-60	3
61-80	8
81-100	12
101-120	7

24) The calculator screens summarize a data set.

```

1-Var Stats
x̄=73.65217391
Σx=1694
Σx²=138696
Sx=25.16239744
σx=24.60931018
↓n=23
  
```

```

1-Var Stats
↑n=23
minX=0
Q1=73
Med=81
Q3=90
maxX=97
  
```

- Identify the mean and the median.
- Based only on the mean and the median, do you expect that the data set is skewed to the right, symmetric, or skewed to the left? Explain.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- The mean and the median are useful measures of central tendency for both qualitative and quantitative data.
 - True
 - False
- In a symmetric and mound shaped distribution, we expect the values of the mean, median, and mode to differ greatly from one another.
 - True
 - False
- In symmetric distributions, the mean and the median will be approximately equal.
 - True
 - False
- In skewed distributions, the mean is the best measure of the center of the distribution since it is least affected by extreme observations.
 - True
 - False
- In practice, the population mean μ is used to estimate the sample mean \bar{x} .
 - True
 - False

- 30) In general, the sample mean is a better estimator of the population mean for larger sample sizes.
 A) True B) False

2.5 Numerical Measures of Variability

1 Calculate Range, Variance, Standard Deviation

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) Each year advertisers spend billions of dollars purchasing commercial time on network television. In the first 6 months of one year, advertisers spent \$1.1 billion. Who were the largest spenders? In a recent article, the top 10 leading spenders and how much each spent (in million of dollars) were listed:

Company A	\$71.4	Company F	\$27.1
Company B	60.1	Company G	25.1
Company C	56.5	Company H	23.1
Company D	55	Company I	22.9
Company E	28.6	Company J	20.6

Calculate the sample variance.

- A) 372.307 B) 3759.244 C) 2070.113 D) 1854.860
- 2) Calculate the range of the following data set:
 9, 6, 9, 2, 8, 10, 6, 6, 9
 A) 8 B) 10 C) 2 D) 12
- 3) The top speeds for a sample of five new automobiles are listed below. Calculate the standard deviation of the speeds. Round to four decimal places.
 180, 100, 195, 130, 155
 A) 38.1772 B) 243.3465 C) 174.0675 D) 137.11
- 4) The amount spent on textbooks for the fall term was recorded for a sample of five university students - \$400, \$350, \$600, \$525, and \$450. Calculate the value of the sample range for the data.
 A) \$99.37 B) \$98.75 C) \$450 D) \$250
- 5) The amount spent on textbooks for the fall term was recorded for a sample of five university students - \$400, \$350, \$600, \$525, and \$450. Calculate the value of the sample standard deviation for the data.
 A) \$99.37 B) \$98.75 C) \$450 D) \$250

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 6) The ages of five randomly chosen professors are 69, 52, 55, 62, and 49. Calculate the sample variance of these ages.
- 7) The data show the total number of medals (gold, silver, and bronze) won by each country winning at least one gold medal in the Winter Olympics. Find the range, sample variance, and sample standard deviation of the numbers of medals won by these countries.

1	2	3	3	4	9	9	11	11
11	14	14	19	22	23	24	25	29

8) The calculator screens summarize a data set.

```

1-Var Stats
x̄=79.95238095
Σx=1679
Σx²=138471
Sx=14.54467666
σx=14.19415101
↓n=21
    
```

```

1-Var Stats
↑n=21
minX=30
Q1=75
Med=82
Q3=90
maxX=97
    
```

- Identify the smallest measurement in the data set.
- Identify the largest measurement in the data set.
- Calculate the range of the data set.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

- Calculate the variance of a sample for which $n = 5$, $\sum x^2 = 1320$, $\sum x = 80$.
 A) 10.00 B) 8.00 C) 326.00 D) 3.16
- Calculate the standard deviation of a sample for which $n = 6$, $\sum x^2 = 830$, $\sum x = 60$.
 A) 6.78 B) 46.00 C) 6.19 D) 164.00
- Compute s^2 and s for the data set: -2, -1, -3, -2, 1, -4
 A) 2.97; 1.72 B) 1.8; 1.34 C) 23.03; 4.8 D) 2.16; 1.47
- Compute s^2 and s for the data set: $\frac{1}{5}, \frac{3}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{10}$.
 A) 0.031; 0.176 B) 0.013; 0.115 C) 0.432; 0.658 D) 3.1; 1.761

2 Interpret Measures of Variability

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- The range of scores on a statistics test was 42. The lowest score was 57. What was the highest score?
 A) 99 B) 78
 C) 70.5 D) cannot be determined
- The temperature fluctuated between a low of 73°F and a high of 89°F. Which of the following could be calculated using just this information?
 A) range B) variance
 C) standard deviation D) median
- Which of the following is a measure of the variability of a distribution?
 A) range B) skewness C) median D) sample size

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- Various state and national automobile associations regularly survey gasoline stations to determine the current retail price of gasoline. Suppose one such national association contacts 200 stations in the United States to determine the price of regular unleaded gasoline at each station. In the context of this problem, define the following descriptive measures: μ , σ , \bar{x} , s .
- Given the sample variance of a distribution, explain how to find the standard deviation.

- 18) Which is expressed in the same units as the original data, the variance or the standard deviation?
- 19) Which measures variability about the mean, the range or the standard deviation?
- 20) For a given data set, which is typically greater, the range or the standard deviation?

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

- 21) The total points scored by a basketball team for each game during its last season have been summarized in the table below. Which statement following the table must be true?

Score	Frequency
41-60	3
61-80	8
81-100	12
101-120	7

- A) The range is at least 41 but at most 79.
 B) The range is 79.
 C) The range is at least 41 but at most 120.
 D) The range is at least 81 but at most 100.
- 22) Which number on the screen below is the sample standard deviation of the data?

```

1-Var Stats
Mx=5.8
Mx=5.8
Mx=408
Sx=2.82055944
σx=2.675817632
n=10
  
```

- A) 2.82 B) 408 C) 5.8 D) 2.67

Answer the question True or False.

- 23) The range is an insensitive measure of data variation for large data sets because two data sets can have the same range but be vastly different with respect to data variation.
 A) True B) False
- 24) For any quantitative data set, $\sum(x - \bar{x}) = 0$.
 A) True B) False
- 25) The sample variance and standard deviation can be calculated using only the sum of the data, $\sum x$, and the sample size, n .
 A) True B) False
- 26) The sample variance is always greater than the sample standard deviation.
 A) True B) False
- 27) A larger standard deviation means greater variability in the data.
 A) True B) False

2.6 Interpreting the Standard Deviation

1 Construct, Interpret Intervals About the Mean

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) The mean \bar{x} of a data set is 36.71, and the sample standard deviation s is 3.22. Find the interval representing measurements within one standard deviation of the mean.
- A) (33.49, 39.93) B) (30.27, 43.15) C) (27.05, 46.37) D) (35.71, 37.71)

- 2) The following is a list of 25 measurements:

12 18 14 17 19 16 14 18 15 17 11
13 14 11 16 18 15 13 17 15 14 19
12 16 17

How many of the measurements fall within one standard deviation of the mean?

- A) 16 B) 18 C) 13 D) 25
- 3) A standardized test has a mean score of 500 points with a standard deviation of 100 points. Five students' scores are shown below.

Adam: 575 Beth: 690 Carlos: 750 Doug: 280 Ella: 440

Which of the students have scores within two standard deviations of the mean?

- A) Adam, Beth, Ella B) Adam, Beth
C) Adam, Beth, Carlos, Ella D) Carlos, Doug

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 4) The mean \bar{x} of a data set is 18, and the sample standard deviation s is 2. Explain what the interval (12, 24) represents.
- 5) The calculator screens summarize a data set.

```
1-Var Stats
x̄=5.5
Σx=55
Σx²=385
Sx=3.027650354
σx=2.872281323
↓n=10
```

```
1-Var Stats
↑n=10
minX=1
Q1=3
Med=5.5
Q3=8
maxX=10
```

- a. Identify the mean and the sample standard deviation. Round to one place after the decimal, where necessary.
- b. Find the interval that corresponds to measurements within two standard deviations of the mean.

- 25) By law, a box of cereal labeled as containing 18 ounces must contain at least 18 ounces of cereal. The machine filling the boxes produces a distribution of fill weights with a mean equal to the setting on the machine and with a standard deviation equal to 0.03 ounce. To ensure that most of the boxes contain at least 18 ounces, the machine is set so that the mean fill per box is 18.09 ounces. Assuming nothing is known about the shape of the distribution, what can be said about the proportion of cereal boxes that contain less than 18 ounces.
- A) The proportion is at most 11%.
 B) The proportion is at least 89%.
 C) The proportion is at most 5.5%.
 D) The proportion is less than 2.5%.
- 26) A study was designed to investigate the effects of two variables — (1) a student's level of mathematical anxiety and (2) teaching method — on a student's achievement in a mathematics course. Students who had a low level of mathematical anxiety were taught using the traditional expository method. These students obtained a mean score of 370 with a standard deviation of 20 on a standardized test. Assuming no information concerning the shape of the distribution is known, what percentage of the students scored between 330 and 410?
- A) at least 75%
 B) approximately 95%
 C) at least 89%
 D) approximately 68%
- 27) A study was designed to investigate the effects of two variables — (1) a student's level of mathematical anxiety and (2) teaching method — on a student's achievement in a mathematics course. Students who had a low level of mathematical anxiety were taught using the traditional expository method. These students obtained a mean score of 470 with a standard deviation of 50 on a standardized test. Assuming a non-mound-shaped distribution, what percentage of the students scored over 620?
- A) at most 11%
 B) approximately 2.5%
 C) at least 89%
 D) at most 5.5%
- 28) A recent survey was conducted to compare the cost of solar energy to the cost of gas or electric energy. Results of the survey revealed that the distribution of the amount of the monthly utility bill of a 3-bedroom house using gas or electric energy had a mean of \$104 and a standard deviation of \$10. If nothing is known about the shape of the distribution, what percentage of homes will have a monthly utility bill of less than \$84?
- A) at most 25%
 B) at least 75%
 C) at most 11.1%
 D) at least 88.9%
- 29) Many firms use on-the-job training to teach their employees computer programming. Suppose you work in the personnel department of a firm that just finished training a group of its employees to program, and you have been requested to review the performance of one of the trainees on the final test that was given to all trainees. The mean and standard deviation of the test scores are 74 and 2, respectively. Assuming nothing is known about the distribution, what percentage of test-takers scored above 80?
- A) at most 11%
 B) at least 89%
 C) approximately 0.15%
 D) approximately 99.85%
- 30) If nothing is known about the shape of a distribution, what percentage of the observations fall within 2 standard deviations of the mean?
- A) at least 75%
 B) at most 25%
 C) approximately 95%
 D) approximately 5%
- 31) Fill in the blank. _____ gives us a method of interpreting the standard deviation of any data set, regardless of the shape of the distribution.
- A) Chebyshev's Rule
 B) The Empirical Rule
 C) both A and B
 D) neither A nor B
- 32) Fill in the blank. _____ is a method of interpreting the standard deviation of data that have a mound-shaped, symmetric distribution.
- A) The Empirical Rule
 B) Chebyshev's Rule
 C) both A and B
 D) neither A nor B

- 33) Given a data set, which of the following is most likely to be the percentage of data within three standard deviations of the mean?
A) 95% B) 65% C) 70% D) 85%

Answer the question True or False.

- 34) Both Chebyshev's rule and the empirical rule guarantee that no data item will be more than four standard deviations from the mean.
A) True B) False
- 35) Chebyshev's rule applies to qualitative data sets, while the empirical rule applies to quantitative data sets.
A) True B) False
- 36) Chebyshev's rule applies to large data sets, while the empirical rule applies to small data sets.
A) True B) False
- 37) Your teacher announces that the scores on a test have a mean of 83 points with a standard deviation of 4 points, so it is reasonable to expect that you scored at least 70 on the test.
A) True B) False

2.7 Numerical Measures of Relative Standing

1 Compute, Interpret z-Score

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) Many firms use on-the-job training to teach their employees computer programming. Suppose you work in the personnel department of a firm that just finished training a group of its employees to program, and you have been requested to review the performance of one of the trainees on the final test that was given to all trainees. The mean and standard deviation of the test scores are 72 and 3, respectively, and the distribution of scores is mound-shaped and symmetric. Suppose the trainee in question received a score of 64. Compute the trainee's z-score.
A) $z = -2.67$ B) $z = -8$ C) $z = -24$ D) $z = 0.85$
- 2) The amount spent on textbooks for the fall term was recorded for a sample of five hundred university students. The mean expenditure was calculated to be \$500 and the standard deviation of the expenditures was calculated to be \$100. Suppose a randomly selected student reported that their textbook expenditure was \$700. Calculate the z-score for this student's textbook expenditure.
A) +2 B) -2 C) +3 D) -3
- 3) A recent survey was conducted to compare the cost of solar energy to the cost of gas or electric energy. Results of the survey revealed that the distribution of the amount of the monthly utility bill of a 3-bedroom house using gas or electric energy had a mean of \$112 and a standard deviation of \$15. Three solar homes reported monthly utility bills of \$59, \$63, and \$64. Which of the following statements is true?
A) Homes using solar power may have lower utility bills than homes using only gas and electricity.
B) The utility bills for homes using solar power are about the same as those for homes using only gas and electricity.
C) Homes using solar power may actually have higher utility bills than homes using only gas and electricity.
D) Homes using solar power always have lower utility bills than homes using only gas and electricity.
- 4) A radio station claims that the amount of advertising each hour has a mean of 16 minutes and a standard deviation of 2.7 minutes. You listen to the radio station for 1 hour and observe that the amount of advertising time is 8 minutes. Calculate the z-score for this amount of advertising time.
A) $z = -2.96$ B) $z = 2.96$ C) $z = -21.6$ D) $z = 0.33$

- 5) On a given day, the price of a gallon of milk had a mean price of \$2.30 with a standard deviation of \$0.05. A particular food store sold milk for \$2.25/gallon. Interpret the z -score for this gas station.
- A) The milk price of this food store falls 1 standard deviation below the milk gas price of all food stores.
 - B) The milk price of this food store falls 1 standard deviation above the mean milk price of all food stores.
 - C) The milk price of this food store falls 5 standard deviations below the mean milk price of all food stores.
 - D) The milk price of this food store falls 5 standard deviations above the mean milk price of all food stores.
- 6) Which of the following is a measure of relative standing?
- A) z -score
 - B) mean
 - C) variance
 - D) pie chart

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 7) A study was designed to investigate the effects of two variables — (1) a student's level of mathematical anxiety and (2) teaching method — on a student's achievement in a mathematics course. Students who had a low level of mathematical anxiety were taught using the traditional expository method. These students obtained a mean score of 380 and a standard deviation of 50 on a standardized test. Find and interpret the z -score of a student who scored 450 on the standardized test.
- 8) A recent survey was conducted to compare the cost of solar energy to the cost of gas or electric energy. Results of the survey revealed that the distribution of the amount of the monthly utility bill of a 3-bedroom house using gas or electric energy had a mean of \$142.00 and a standard deviation of \$15.00. Assuming the distribution is mound-shaped and symmetric, would you expect to see a 3-bedroom house using gas or electric energy with a monthly utility bill of \$239.50? Explain.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

- 9) Find the z -score for the value 100, when the mean is 69 and the standard deviation is 5.
- A) $z = 6.20$
 - B) $z = 6.00$
 - C) $z = -1.38$
 - D) $z = 1.38$

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 10) Test scores for a history class had a mean of 79 with a standard deviation of 4.5. Test scores for a physics class had a mean of 69 with a standard deviation of 3.7. One student earned a 86 on the history test and a 85 on the physics test. Calculate the z -score for each test. On which test did the student perform better?
- 11) The following data represent the scores of 50 students on a statistics exam. The mean score is 80.02, and the standard deviation is 11.9.

39 51 59 63 66 68 68 69 70 71
 71 71 73 74 76 76 76 77 78 79
 79 79 79 80 80 82 83 83 83 85
 85 86 86 88 88 88 88 89 89 89
 90 90 91 91 92 95 96 97 97 98

Find the z -scores for the highest and lowest exam scores.

- 12) The z -score for a value x is -2.5 . State whether the value of x lies above or below the mean and by how many standard deviations.
- 13) Suppose that 50 and 75 are two elements of a population data set and their z -scores are -3 and 2 , respectively. Find the mean and standard deviation.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- 14) According to the empirical rule, z -scores of less than -3 or greater than 3 occur very infrequently for data from a mound and symmetric distribution
A) True B) False
- 15) If a z -score is 0 or near 0 , the measurement is located at or near the mean.
A) True B) False
- 16) If a sample has mean 0 and standard deviation 1 , then for every measurement x in the sample the z -score of x is x itself.
A) True B) False

2 Find, Interpret Percentile

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 17) When Scholastic Achievement Test scores (SATs) are sent to test-takers, the percentiles associated with scores are also given. Suppose a test-taker scored at the 67th percentile on the verbal part of the test and at the 38th percentile on the quantitative part. Interpret these results.
A) This student performed better than 67% of the other test-takers on the verbal part and better than 38% on the quantitative part.
B) This student performed better than 67% of the other test-takers on the verbal part and better than 62% on the quantitative part.
C) This student performed better than 33% of the other test-takers on the verbal part and better than 62% on the quantitative part.
D) This student performed better than 33% of the other test-takers on the verbal part and better than 38% on the quantitative part.
- 18) The amount spent on textbooks for the fall term was recorded for a sample of five hundred university students. It was determined that the 75th percentile was the value \$500. Which of the following interpretations of the 75th percentile is correct?
A) The average of the 500 textbook costs was \$500.
B) 75% of the students sampled had textbook costs that exceeded \$500.
C) 75% of the students sampled had textbook costs equal to \$500.
D) 25% of the students sampled had textbook costs that exceeded \$500.
- 19) Summary information is given for the weights (in pounds) of 1000 randomly sampled tractor trailers.

MIN:	4015	25%:	5615
MAX:	10,615	75%:	8615
AVE:	7015	Std. Dev.:	1400

Find the percentage of tractor trailers with weights between 5615 and 8615 pounds.

- A) 50% B) 75% C) 25% D) 100%
- 20) The test scores of 30 students are listed below. Which number could be the 30th percentile?
31 41 45 48 52 55 56 56 63 65
67 67 69 70 70 74 75 78 79 79
80 81 83 85 85 87 90 92 95 99
A) 64 B) 56 C) 67 D) 90

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 21) A retail store's customer satisfaction rating is at the 88th percentile. What percentage of retail stores has higher customer satisfaction ratings than this store?
- 22) In a summary of recent real estate sales, the median home price is given as \$325,000. What percentile corresponds to a home price of \$325,000?

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- 23) The mean of a data set is at the 50th percentile.
A) True B) False
- 24) Percentile rankings are of practical value only with large data sets.
A) True B) False
- 25) The process for finding a percentile is similar to the process for finding the median.
A) True B) False

2.8 Methods for Detecting Outliers: Box Plots and z-Scores

1 Determine if Datum is an Outlier

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) At the U.S. Open Tennis Championship a statistician keeps track of every serve that a player hits during the tournament. The statistician reported that the mean serve speed of a particular player was 103 miles per hour (mph) and the standard deviation of the serve speeds was 12 mph. Using the z-score approach for detecting outliers, which of the following serve speeds would represent outliers in the distribution of the player's serve speeds?

Speeds: 61 mph, 115 mph, and 127 mph

- A) 61 is the only outlier. B) 61 and 115 are both outliers, but 127 is not.
C) 61, 115, and 127 are all outliers. D) None of the three speeds is an outlier.
- 2) At the U.S. Open Tennis Championship a statistician keeps track of every serve that a player hits during the tournament. The statistician reported that the mean serve speed of a particular player was 100 miles per hour (mph) and the standard deviation of the serve speeds was 15 mph. Using the z-score approach for detecting outliers, which of the following serve speeds would represent outliers in the distribution of the player's serve speeds?

Speeds: 50 mph, 80 mph, and 105 mph

- A) 50 is the only outlier. B) 50 and 80 are both outliers, 105 is not.
C) 50, 80, and 105 are all outliers. D) None of the three speeds are outliers.
- 3) The speeds of the fastballs thrown by major league baseball pitchers were measured by radar gun. The mean speed was 87 miles per hour. The standard deviation of the speeds was 3 mph. Which of the following speeds would be classified as an outlier?

A) 97 mph B) 81 mph C) 92 mph D) 84 mph

- 4) Which of the following statements concerning the box plot and z-score methods for detecting outliers is false?
- A) The z-score method is less affected by an extreme observation in the data set.
 - B) The box plot method uses the quartiles as a basis for detecting outliers.
 - C) The z-score method uses the mean and standard deviation as a basis for detecting outliers.
 - D) The box plot method is less affected by an extreme observation in the data set.
- 5) Which of the following statements could be an explanation for the presence of an outlier in the data?
- A) The measurement may be correct and from the same population as the rest but represents a rare event. Generally, we accept this explanation only after carefully ruling out all others.
 - B) The measurement is incorrect. It may have been observed, recorded, or entered into the computer incorrectly.
 - C) The measurement belongs to a population different from that from which the rest of the sample was drawn.
 - D) All of the above are explanations for outliers.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 6) A radio station claims that the amount of advertising each hour has an a mean of 15 minutes and a standard deviation of 1.2 minutes. You listen to the radio station for 1 hour and observe that the amount of advertising time is 12.96 minutes. Based on your observation, what would you infer about the radio station's claim?
- 7) The following data represent the scores of 50 students on a statistics exam. The mean score is 80.02, and the standard deviation is 11.9.

39	51	59	63	66	68	68	69	70	71
71	71	73	74	76	76	76	77	78	79
79	79	79	80	80	82	83	83	83	85
85	86	86	88	88	88	88	89	89	89
90	90	91	91	92	95	96	97	97	98

Use the z-score method to identify potential outliers among the scores.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

- 8) The z-score uses the quartiles to identify outliers in a data set.
- A) True
 - B) False
- 9) An outlier is defined as any observation that falls within the outer fences of a box plot.
- A) True
 - B) False
- 10) Box plots are used to detect outliers in qualitative data sets, while z-scores are used to detect outliers in quantitative data sets.
- A) True
 - B) False
- 11) An outlier in a data set may have a simple explanation such as a scale was not working properly or the researcher inverted the digits of a number when recording a measurement.
- A) True
 - B) False
- 12) An outlier may be caused by accidentally including the height of a six-year-old boy in a set of data representing the heights of 12-year-old boys.
- A) True
 - B) False
- 13) The outer fences of a box plot are three standard deviations from the mean.
- A) True
 - B) False

2 Calculate Quartiles and IQR

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 14) At the U.S. Open Tennis Championship a statistician keeps track of every serve that a player hits during the tournament. The lower quartile of a particular player's serve speeds was reported to be 94 mph. Which of the following interpretations of this information is correct?
- A) 75% of the player's serves were hit at speeds greater than 94 mph.
 - B) 25% of the player's serves were hit at 94 mph.
 - C) 75% of the player's serves were hit at speeds less than 94 mph.
 - D) 94 serves traveled faster than the lower quartile.
- 15) A sociologist recently conducted a survey of citizens over 60 years of age who have net worths too high to qualify for Medicaid but have no private health insurance. The ages of the 25 uninsured senior citizens were as follows:

68 73 66 76 86 74 61 89 65 90 69 92 76
62 81 63 68 81 70 73 60 87 75 64 82

Find the upper quartile of the data.

- A) 81.5 B) 65.5 C) 73 D) 92

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 16) The amount of television viewed by today's youth is of primary concern to Parents Against Watching Television (PAWT). Three hundred parents of elementary school-aged children were asked to estimate the number of hours per week that their child watches television. The upper quartile for the distribution was given as 15 hours. Interpret this value.
- 17) For a given data set, the lower quartile is 45, the median is 50, and the upper quartile is 57. The minimum value in the data set is 32, and the maximum is 81.
- a. Find the interquartile range.
 - b. Find the inner fences.
 - c. Find the outer fences.
 - d. Is either of the minimum or maximum values considered an outlier? Explain.
- 18) The calculator screens summarize a data set.

```
1-Var Stats
x̄=79.95238095
Σx=1679
Σx²=138471
Sx=14.54467666
σx=14.19415101
↓n=21
```

```
1-Var Stats
↑n=21
minX=30
Q1=75
Med=82
Q3=90
maxX=97
```

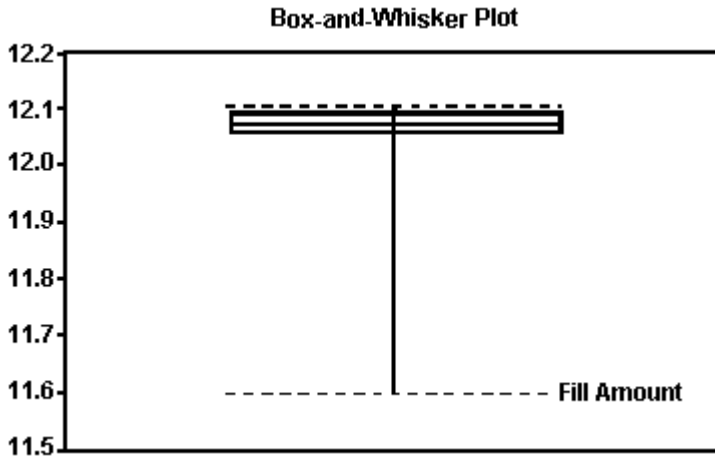
- a. Identify the lower and upper quartiles of the data set.
- b. Find the interquartile range.
- c. Is there reason to suspect that the data may contain an outlier? Explain.

3 Construct, Interpret Boxplot

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

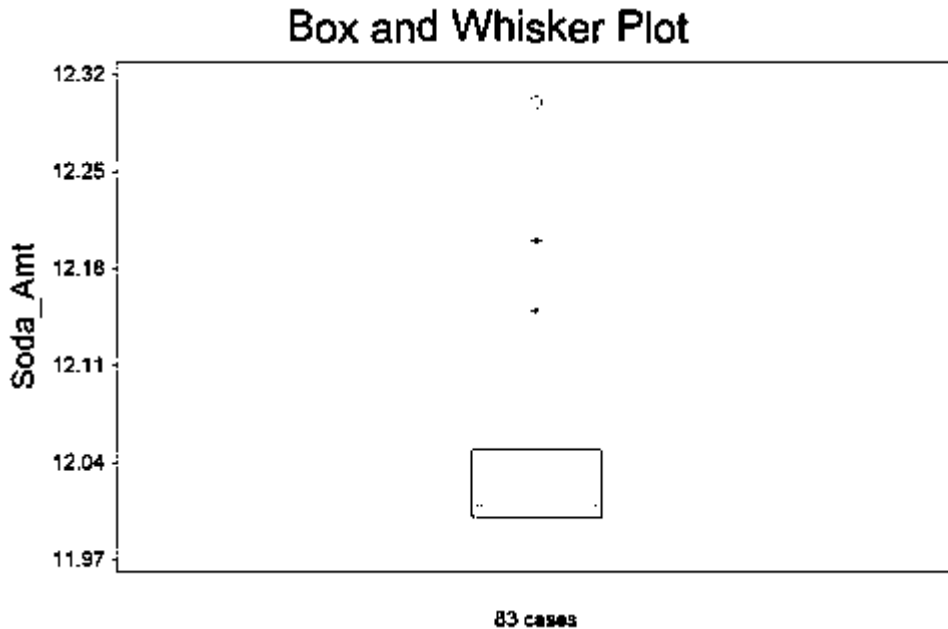
Solve the problem.

- 19) The box plot shown below displays the amount of soda that was poured by a filling machine into 12 -ounce soda cans at a local bottling company.



Based on the box plot, what shape do you believe the distribution of the data to have?

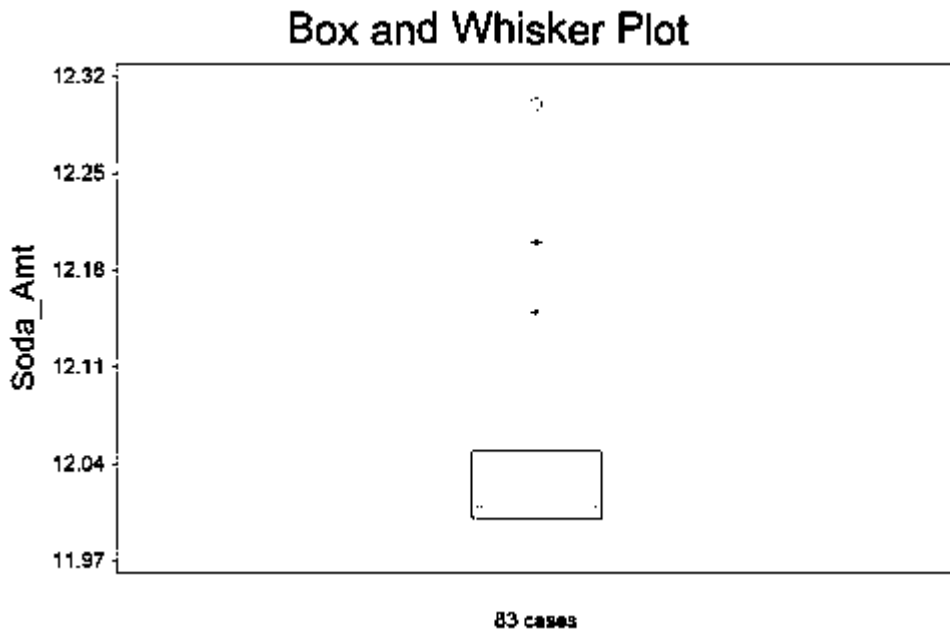
- A) skewed to the left
B) approximately symmetric
C) skewed to the right
D) skewed to the center
- 20) The box plot shown below was constructed for the amount of soda that was poured by a filling machine into 12-ounce soda cans at a local soda bottling company.



We see that one soda can received 12.15 ounces of soda on the plot above. Based on the box plot presented, how would you classify this observation?

- A) expected observation
B) suspect outlier
C) highly suspect outlier
D) it has a lot of soda

21) The box plot shown below was constructed for the amount of soda that was poured by a filling machine into 12-ounce soda cans at a local soda bottling company.



We see that one soda can received 12.30 ounces of soda on the plot above. Based on the box plot presented, how would you classify this observation?

- A) expected observation
- B) suspect outlier
- C) highly suspect outlier
- D) it has a lot of soda

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

22) The following data represent the scores of 50 students on a statistics exam.

39	51	59	63	66	68	68	69	70	71
71	71	73	74	76	76	76	77	78	79
79	79	79	80	80	82	83	83	83	85
85	86	86	88	88	88	88	89	89	89
90	90	91	91	92	95	96	97	97	98

- a. Find the lower quartile, the upper quartile, and the median of the scores.
- b. Find the interquartile range of the data and use it to identify potential outliers.
- c. In a box plot for the data, which scores, if any, would be outside the outer fences? Which scores, if any, would be outside the inner fences but inside the outer fences?

23) Use a graphing calculator or software to construct a box plot for the following data set.

12	18	14	17	19	16	14	18	15	17	11
13	14	11	16	18	15	13	17	15	14	19
12	16	17								

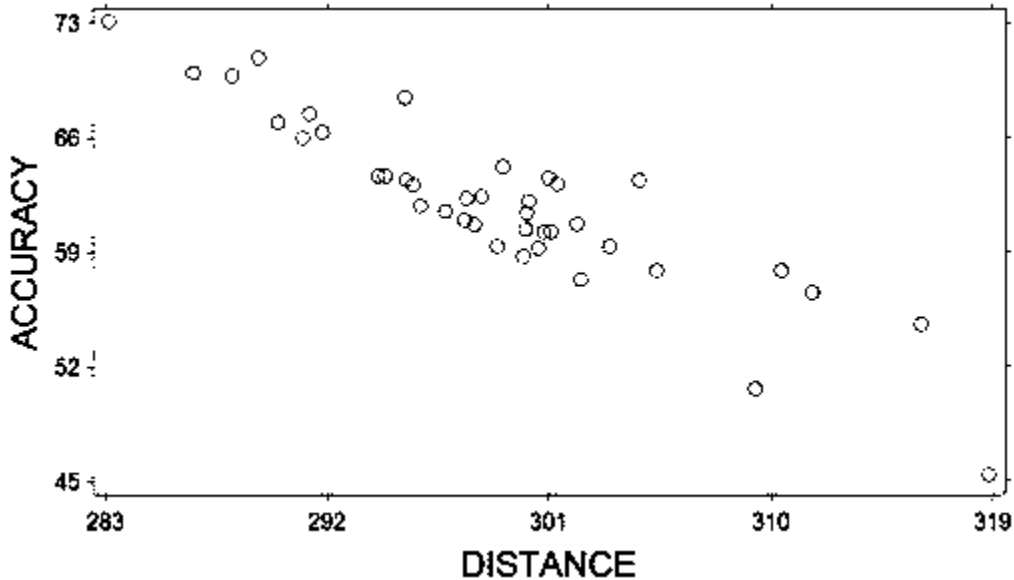
2.9 Graphing Bivariate Relationships (Optional)

1 Construct, Interpret Scatterplot

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

- 1) A sample of professional golfers was taken and their driving distance (measured as the average distance as their drive off the tee) and driving accuracy (measured as the percentage of fairways that their drives landed in) were recorded. A scatterplot of the variables is shown below.



What relationship do these two variables exhibit?

- A) They exhibit no relationship
B) They exhibit a positive linear relationship
C) They exhibit a negative linear relationship
D) They exhibit a curvilinear relationship

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

- 2) The data below represent the numbers of absences and the final grades of 15 randomly selected students from a statistics class. Construct a scattergram for the data. Do you detect a trend?

Student	Number of Absences	Final Grade as a Percent
1	5	79
2	6	78
3	2	86
4	12	56
5	9	75
6	5	90
7	8	78
8	15	48
9	0	92
10	1	78
11	9	81
12	3	86
13	10	75
14	3	89
15	11	65

3) The scores of nine members of a women's golf team in two rounds of tournament play are listed below.

Player	1	2	3	4	5	6	7	8	9
Round 1	85	90	87	78	92	85	79	93	86
Round 2	90	87	85	84	86	78	77	91	82

Construct a scattergram for the data.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

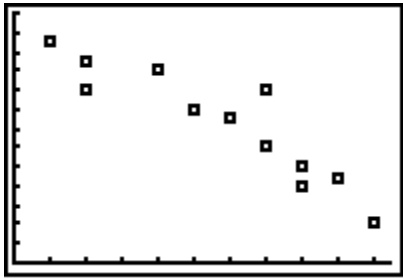
Answer the question True or False.

4) Scatterplots are useful for both qualitative and quantitative data.

A) True

B) False

5) The scatterplot below shows a negative relationship between two variables.



A) True

B) False

2.10 Distorting the Truth with Descriptive Statistics

1 Understand Misleading Statistics

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

1) Explain how stretching the vertical axis of a histogram can be misleading.

2) Explain how using a scale break on the vertical axis of a histogram can be misleading.

3) Explain how it can be misleading to draw the bars in a histogram so that the width of each bar is proportional to its height rather than have all bars the same width.

4) Explain how it can be misleading to report only the mean of a distribution without any measure of the variability.

Ch. 2 Methods for Describing Sets of Data

Answer Key

2.1 Describing Qualitative Data

1 Identify Classes/Compute Class Frequencies/Relative Frequencies/Percentages

- 1) A
- 2) A
- 3) free account, institutional account, account paid for personally

2 Construct Frequency/Relative Frequency Table

- 4) A
- 5) A
- 6)

Color	Frequency
Green	3
Blue	7
Brown	5
Orange	2
Red	3

- 7) B

3 Construct, Interpret Bar Graph

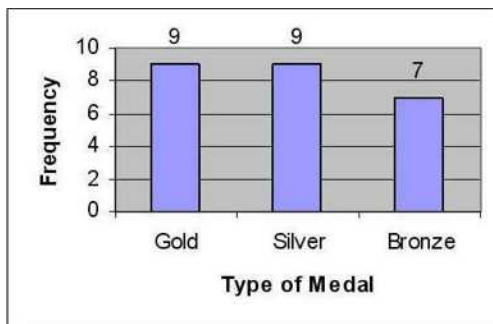
- 8) B
- 9) A
- 10) A
- 11) a.

Medal	Frequency
Gold	9
Silver	9
Bronze	7

- b.

Medal	Relative Frequency
Gold	.36
Silver	.36
Bronze	.28

- c.



- 12) A
- 13) A

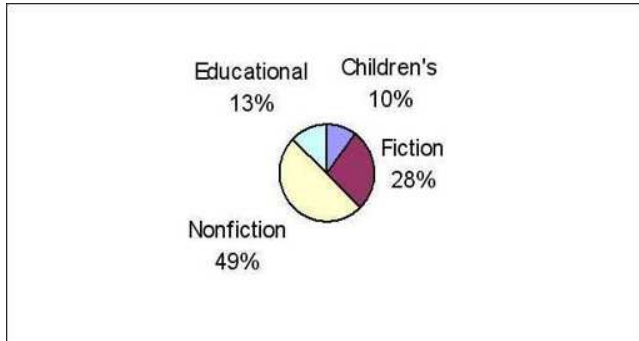
4 Construct, Interpret Pie Chart

- 14) A
- 15) C

16) a.

Type of Book	Relative Frequency
Children's	.10
Fiction	.28
Nonfiction	.49
Educational	.13

b.



17) A

18) B

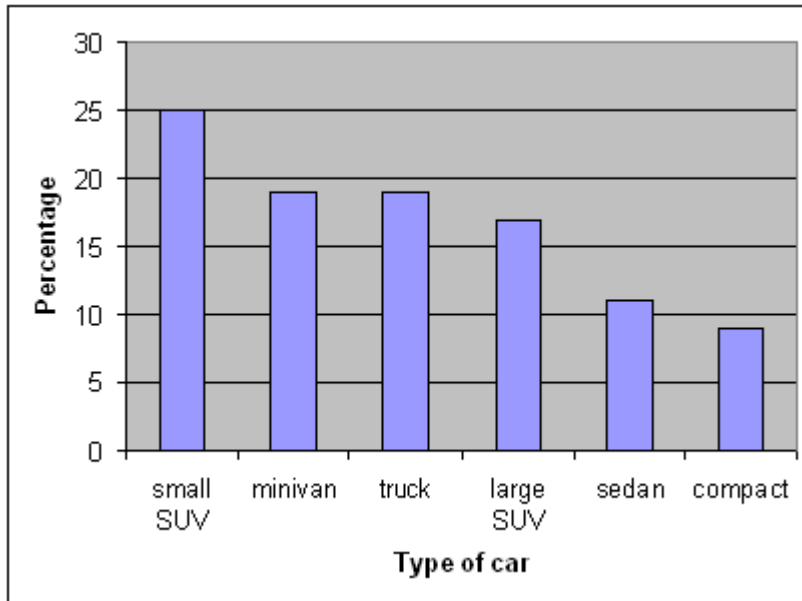
5 Construct, Interpret Pareto Diagram

19) In a Pareto diagram, the bars are arranged by height in a descending order from left to right.

20) a.

Car	Relative Frequency
compact	0.09
sedan	0.11
small SUV	0.25
large SUV	0.17
minivan	0.19
truck	0.19

b.

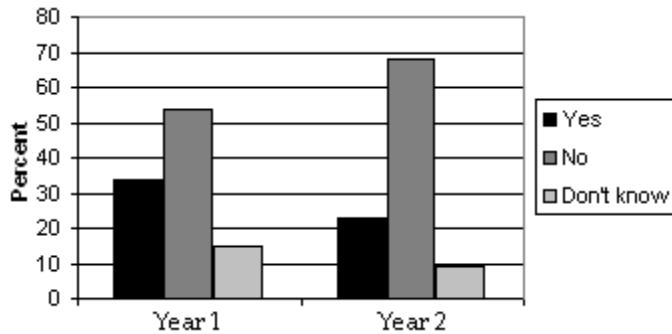


21) B

22) B

6 Construct, Interpret Side-by-Side Bar Chart

23)



Losses due to employee theft have decreased from year 1 to year 2.

2.2 Graphical Methods for Describing Quantitative Data

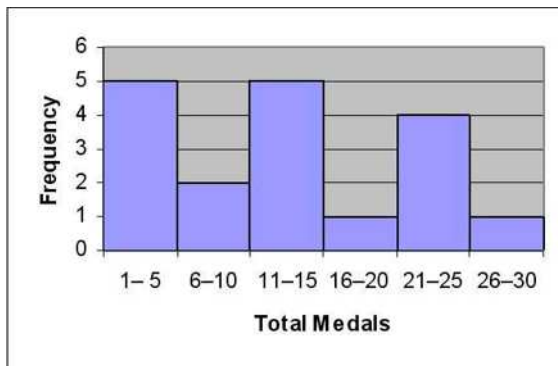
1 Construct, Interpret Histogram

1) B

2) a.

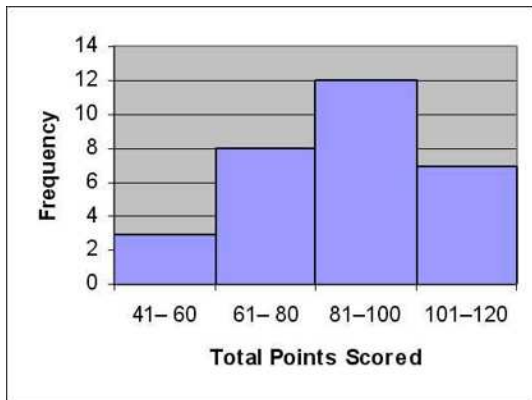
Total Medals	Frequency
1-5	5
6-10	2
11-15	5
16-20	1
21-25	4
26-30	1

b.



3) a. The exact scores would be needed to construct a stem-and-leaf display but the exact scores are not available in the table given.

b.



4) A

5) A

6) B

2 Construct, Interpret Stem-and-Leaf Display

7) A

8) C

9)

Stem	Leaf
1	8
2	
3	
4	
5	4 6
6	6 9
7	6 7 9
8	2 3 5 7 8 8 9
9	1 4 6 6 8

10) B

3 Construct, Interpret Dot-Plot

11) A

12) A

13) A

14) A

2.3 Summation Notation

1 Use Summation Notation

1) A

2) A

3) A

4) A

5) A

6) A

2.4 Numerical Measures of Central Tendency

1 Find Mean, Median, Mode

1) B

2) A

3) A

4) A

5) A

6) A

6) The mean of the data is $x = \frac{\sum x}{n}$

$$\frac{70.4 + 61.7 + 57.2 + 54.5 + 31.4 + 24.6 + 24.3 + 23.1 + 21.8 + 19.9}{10}$$
$$= \frac{388.9}{10}$$
$$= 38.89 \Rightarrow \$38.89 \text{ million}$$

The median is the average of the middle two observations.

$$M = \frac{31.4 + 24.6}{2} = 28.00 \Rightarrow \$28.00 \text{ million}$$

7) The mean is the sum of the numbers divided by 18:

$$\frac{1 + 2 + 3 + 3 + 4 + 9 + 9 + 11 + 11 + 11 + 14 + 14 + 19 + 22 + 23 + 24 + 25 + 29}{18}$$
$$= \frac{234}{18} = 13 \text{ medals.}$$

The median is the mean of the two middle numbers: $\frac{11 + 11}{2} = 11$ medals.

The mode is the most frequent number of medals: 11 medals.

8) The mean is divided by n:

$$\frac{\sum x}{n} = \frac{196}{8} = 24.5.$$

9) a. $n = 21$

b. $\sum x = 1679$

c. mean: $\bar{x} \approx 79.95$; median: Med=82; mode: not possible

2 Interpret Measures of Central Tendency

10) A

11) A

- 12) D
 13) A
 14) A
 15) A
 16) A
 17) A
 18) A
 19) A
 20) A
 21) Since the distribution is skewed to the left, we know that the median time will exceed the mean time.
 22) In both year 1 and year 5, the mean dropout rates exceed the median dropout rates. This indicates that both the year 1 and year 5 high school dropout rates have distributions that are skewed to the right.
 23) The modal class is the class with the greatest frequency: 81–100 points.
 24) a. mean: $\bar{x} \approx 73.65$; median: Med=81
 b. We expect the data to be skewed to the left because the mean is less than the median.
 25) B
 26) B
 27) A
 28) B
 29) B
 30) A

2.5 Numerical Measures of Variability

1 Calculate Range, Variance, Standard Deviation

- 1) A
 2) A
 3) A
 4) D
 5) A

$$6) s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{69 + 52 + 55 + 62 + 49}{5} = 57.4$$

$$s^2 = \frac{(69 - 57.4)^2 + (52 - 57.4)^2 + (55 - 57.4)^2 + (62 - 57.4)^2 + (49 - 57.4)^2}{5 - 1}$$

$$= 65.30$$

- 7) The range is $29 - 1 = 28$ medals.

$$\text{The variance is } s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1} = \frac{4372 - \frac{(234)^2}{18}}{17} = \frac{1330}{17} \approx 78.24$$

$$\text{The standard deviation is } s = \sqrt{s^2} = \sqrt{\frac{1330}{17}} \approx 8.85$$

- 8) a. minX=30
 b. maxX=97
 c. $97 - 30 = 67$
 9) A
 10) A
 11) A
 12) A

2 Interpret Measures of Variability

- 13) A
- 14) A
- 15) A
- 16) μ is the mean price of the regular unleaded gasoline prices of all retail gas stations in the United States.

σ is the standard deviation of the regular unleaded gasoline prices of all retail gas stations in the United States.

\bar{x} is the mean price of the regular unleaded gasoline prices collected from the 200 stations sampled.

s is the standard deviation of the regular unleaded gasoline prices collected from the 200 stations sampled.

17) Take the square root of the sample variance to find the sample standard deviation.

18) standard deviation

19) standard deviation

20) range

21) A

22) A

23) A

24) A

25) B

26) B

27) A

2.6 Interpreting the Standard Deviation

1 Construct, Interpret Intervals About the Mean

- 1) A
- 2) A
- 3) A
- 4) measurements within three standard deviations of the mean
- 5) a. mean: $\bar{x} = 5.5$; sample standard deviation: $S_x \approx 3.0$
 - b. $(5.5 - 2 \times 3.0, 5.5 + 2 \times 3.0) = (-.5, 11.5)$

2 Use Empirical Rule

- 6) A
- 7) A
- 8) D
- 9) A
- 10) A
- 11) A
- 12) A
- 13) A
- 14) A
- 15) A
- 16) We use the Empirical Rule to determine the percentage of serves with speeds faster than 82 mph. We do this by first finding the percentage of serves with speeds between 82 and 104 mph. The Empirical Rule states that approximately 34.0% (68%/2) fall between 82 and 104 mph. Because the distribution is symmetric about the mean speed of 104 mph, we know 50% of the serve speeds were faster than 104 mph. We add these findings together to determine that $34.0\% + 50\% = 84.0\%$ of the serves were hit faster than 82 mph.
- 17) The value 91 falls one standard deviation above the mean in the distribution. Using the Empirical Rule, 68% of the days will have between 81 and 91 jobs submitted. Of the remaining 32% of the days, half, or $32\%/2 = 16\%$, of the days will have more than 91 jobs submitted.
- 18) The value of 36 ounces falls three standard deviations below the mean. The Empirical Rule states that approximately all of the boxes will contain cereal amounts between 36.00 ounces and 36.24 ounces. Therefore, approximately 100% of the boxes contain at least 36 ounces.

- 19) The Empirical Rule states that 95% of the data will fall between 77 and 89. Because the distribution is symmetric, half of the remaining 5%, or 2.5%, will have test scores above 89. Thus, 89 is the cutoff point that will identify the trainees who will receive the promotion.
- 20) 74% of the scores lie within one standard deviation of the mean, 96% within two standard deviations, and 98% within three standard deviations. These percentages are close to those given in the Empirical Rule, so the distribution is roughly mound-shaped and symmetric, though obviously skewed slightly to the left.
- 21) A

3 Use Chebyshev's Rule

- 22) D
 23) A
 24) B
 25) A
 26) A
 27) A
 28) A
 29) A
 30) A
 31) A
 32) A
 33) A
 34) B
 35) B
 36) B
 37) A

2.7 Numerical Measures of Relative Standing

1 Compute, Interpret z-Score

- 1) A
 2) A
 3) A
 4) A
 5) A
 6) A
 7) The z-score is $z = \frac{x - \mu}{\sigma}$.

For a score of 45, $z = \frac{450 - 380}{50} = 1.40$.

This student's score falls 1.40 standard deviations above the mean score of 380.

- 8) The z-score for the value \$239.50 is:

$$z = \frac{x - \bar{x}}{s} = \frac{239.5 - 142}{15} = 6.5$$

An observation that falls 6.5 standard deviations above the mean is very unlikely. We would not expect to see a monthly utility bill of \$239.50 for this home.

- 9) A
 10) history z-score = 1.56; physics z-score = 4.32; The student performed better on the physics test.
 11) highest: $z = 1.51$; lowest: $z = -3.45$
 12) The value of x lies 2.5 standard deviations below the mean.
 13) mean: 65; standard deviation: 5
 14) A
 15) A
 16) A

2 Find, Interpret Percentile

- 17) A
- 18) D
- 19) A
- 20) A
- 21) 12%
- 22) 50th percentile
- 23) B
- 24) A
- 25) A

2.8 Methods for Detecting Outliers: Box Plots and z-Scores

1 Determine if Datum is an Outlier

- 1) A
- 2) A
- 3) A
- 4) A
- 5) D
- 6) The z-score for the value 12.96 is -1.7 .
Since the z-score would not indicate that 12.96 minutes represents an outlier, there is no evidence that the station's claim is incorrect.
- 7) The z-score of 39 is -3.46 . Since this z-score is less than -3 , the score of 39 is an outlier. All other scores have z-scores between -3 and 3 , so there are no other outliers.
- 8) B
- 9) B
- 10) B
- 11) A
- 12) A
- 13) B

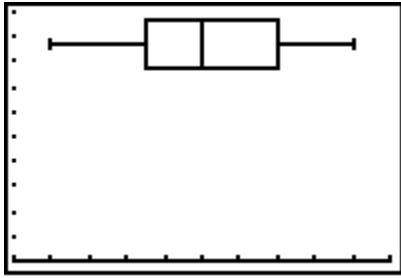
2 Calculate Quartiles and IQR

- 14) A
- 15) A
- 16) 75% of the TV viewing times are less than 15 hours per week. 25% of the times exceed 15 hours per week.
- 17) a. The interquartile range is $57 - 45 = 12$.
b. The inner fences are $45 - 1.5(12) = 27$ and $57 + 1.5(12) = 75$.
c. The outer fences are $45 - 3(12) = 9$ and $57 + 3(12) = 93$.
d. The maximum of 81 is a potential outlier since it lies outside the inner fences. The minimum is within the inner fence and is not considered to be an outlier.
- 18) a. lower quartile: $Q1=75$; upper quartile: $Q3=90$
b. interquartile range: $90 - 75 = 15$
c. Yes; the smallest measurement, 30, is three times the interquartile range less than the lower quartile, so it is a suspected outlier.

3 Construct, Interpret Boxplot

- 19) A
- 20) B
- 21) C
- 22) a. The lower quartile is 73, the upper quartile is 89, and the median is 81.
b. The interquartile range is $89 - 73 = 16$. The score of 39 is a potential outlier since it is less than $73 - 1.5(16) = 49$.
c. No scores fall outside the outer fences, 25 and 137. Only the score of 39 lies between the inner and outer fences.

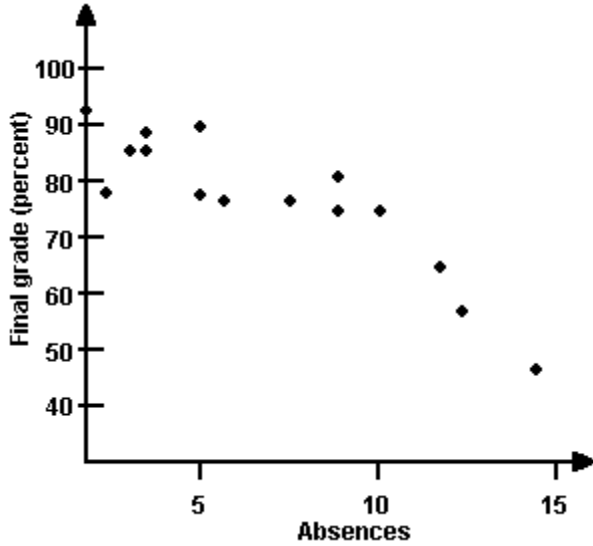
23) The horizontal axis extends from 10 to 20, with each tick mark representing one unit.



2.9 Graphing Bivariate Relationships (Optional)

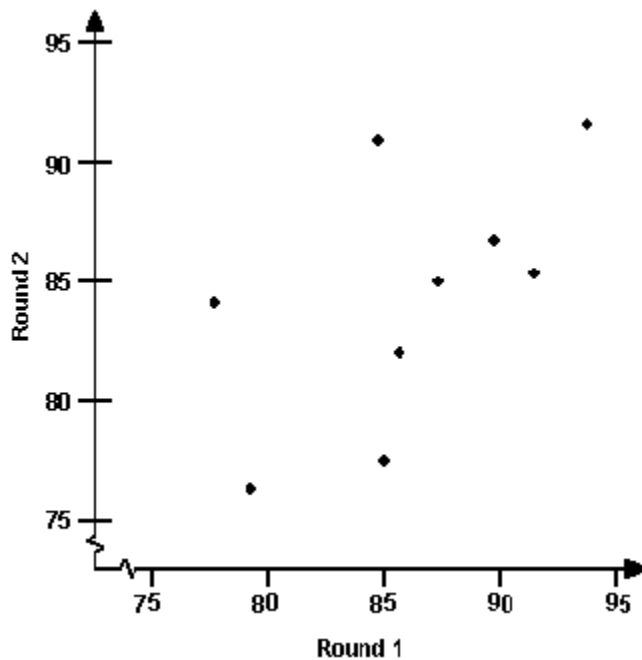
1 Construct, Interpret Scatterplot

- 1) C
- 2)



There appears to be a trend in the data. As the number of absences increases, the final grade decreases.

3)



4) B

5) A

2.10 Distorting the Truth with Descriptive Statistics

1 Understand Misleading Statistics

- 1) Stretching the vertical axis may overemphasize the differences in the heights of the bars making the taller bars look much taller than the shorter bars.
- 2) Using a scale break on the vertical axis may make the shorter bars look disproportionately shorter than the taller bars.
- 3) The reader may think that the area of the bar represents the quantity rather than the height of the bar, giving a disproportionate emphasis on the taller bars.
- 4) When comparing means from two different distributions, the difference between them may be insignificant if the variability in one or both of the distributions is large.