CAMOSUN COLLEGE

HOMEWORK PROBLEMS

MATH 156

Statistics Course Materials

Department of Mathematics & Statistics
Camosun College

ii

Section 5.1: Exercises

1.3 **EXERCISES**

UNDERSTANDING THE CONCEPTS

1.1 **Experimental Units** Identify the experimental units on which the following variables are measured:

a. Gender of a student

b. Number of errors on a midterm exam

c. Age of a cancer patient

d. Number of flowers on an azalea plant

e. Colour of a car entering the parking lot

1.2 **Qualitative or Quantitative?** Identify each variable as quantitative or qualitative:

a. Amount of time it takes to assemble a simple puzzle

b. Number of students in a first-grade classroom

c. Rating of a newly elected politician (excellent, good, fair, poor)

d. Province or territory in which a person lives

1.3 **Discrete or Continuous?** Identify the following quantitative variables as discrete or continuous:

a. Population in a particular area of Canada

b. Weight of newspapers recovered for recycling on a single day

c. Time to complete a sociology exam

d. Number of consumers in a poll of 1000 who consider nutritional labelling on food products to be important

1.4 **Discrete or Continuous?** Identify each quantitative variable as discrete or continuous.

a. Number of boating accidents along a 50-kilometre stretch of the St. Lawrence River

b. Time required to complete a questionnaire

c. Choice of colour for a new refrigerator

d. Number of brothers and sisters you have

e. Yield in kilograms of wheat from a 10,000-square-metre plot in a wheat field

1.5 **Parking on Campus** Six vehicles are selected from the vehicles that are issued campus parking permits, and the following data are recorded:

Vehicle	Type	Make	Carpool?	One-way Commute Distance (kilometres)	Age of Vehicle (years)
1	Car	Honda	No	23.6	6
2	Car	Toyota	No	17.2	3
3	Truck	Toyota	No	10.1	4
4	Van	Dodge	Yes	31.7	2
5	Motor-cycle	Harley-Davidson	No	25.5	1
6	Car	Chevrolet	No	5.4	9

a. What are the experimental units?

b. What are the variables being measured? What types of variables are they?

c. Is this univariate, bivariate, or multivariate data?

1.6 **Past Canadian Prime Ministers** A data set consists of the ages at death for each of the 15 past prime ministers of Canada.

a. Is this set of measurements a population or a sample?

b. What is the variable being measured?

c. Is the variable in part b quantitative or qualitative?

1.7 **Voter Attitudes** You are a candidate for your provincial assembly, and you want to survey voter attitudes regarding your chances of winning. Identify the population that is of interest to you and from which you would like to select your sample. How is this population dependent on time?

1.8 **Cancer Survival Times** A medical researcher wants to estimate the survival time of a patient after the onset of a particular type of cancer and after a particular regimen of radiotherapy.

a. What is the variable of interest to the medical researcher?

b. Is the variable in part a qualitative, quantitative discrete, or quantitative continuous?

c. Identify the population of interest to the medical researcher.

d. Describe how the researcher could select a sample from the population.

e. What problems might arise in sampling from this population?

1.9 **New Teaching Methods** An educational researcher wants to evaluate the effectiveness of a new method for teaching reading to deaf students. Achievement at the end of a period of teaching is measured by a student's score on a reading test.

a. What is the variable to be measured? What type of variable is it?

b. What is the experimental unit?

c. Identify the population of interest to the experimenter.

Section 5.1: Answers

Answers to Selected Exercises

Chapter 1

1.1 a. the student   b. the exam
c. the patient   d. the plant   e. the car

1.3 a. discrete   b. continuous
c. continuous   d. discrete

1.5 a. vehicles   b. type (qualitative); make (qualitative); carpool (qualitative); distance (quantitative continuous); age (quantitative continuous)   c. multivariate

1.7 The population is the set of voter preferences for all voters in the province. Voter preferences may change over time.

1.9 a. score on the reading test; quantitative
b. the student   c. the set of scores for all deaf students who hypothetically might take the test

1.11 a. a pair of jeans   b. the province in which the jeans are produced; qualitative   e. 8/25
f. Ontario   g. The three provinces produce roughly the same numbers of jeans.

1.13 a. no

1.15 a. yes   b. yes   c. the bar chart

1.17 a. Mound-shaped distribution
b. 1.6 (1 6)
c. both 4.9 (4 9)

1.19 a. 3 | 2 3 4 5 5 5 6 6 7 9 9 9 9
4 | 0 0 2 2 3 3 3 4 4 4 5 8   leaf digit = 0.1, 1 2 represents 1.2
b. 3 | 2 3 4
3 | 5 5 5 6 6 7 9 9 9 9
4 | 0 0 2 2 3 3 3 4 4   leaf digit = 0.1, 1 2
4 | 5 8   represents 1.2
Yes.

1.21 b. The ones digit must be the stem, and the leaf will be a zero digit.
c. 0 | 0 0 0 0 0
1 | 0 0 0 0 0 0 0 0 0
2 | 0 0 0 0 0 0
d. Yes, if the stem and leaf plot is turned 90 degrees and stretched to resemble the dotplot.

1.25 b. Skewed right
c. 0.72

1.27 a. 3 | 0 0 0 1 1 2 2 2 3 3 4 4
3 | 5 5 5 6 6 6 6 6 7 7 8 8 9 9 9
4 | 0 0 0 0 1 1 1 1 2 2 3 3
4 | 5 5 6 6 6 7 8 8
5 | 0 0
5 | 5
c. Both are very similar; however, the relative frequency histogram may be more helpful.
d. .54
e. .94

1.29 a. 0 | 2 2 3 3 3 4 4 4
0 | 5 5 6 6 6 6 7 7 7 8 8 8 8 9 9
1 | 0 0 1 1 1 1 1 1 1 2 2 2 3 3 3 4 4
1 | 6 6 7 7 8 8 8 8 9 9
2 | 1 1 2 3
2 | 5 8   leaf digit = 0.1, 1 2 represents 1.2
3 | 1 1
3 | 6
4 |
4 | 5
5 | 2
b. 0.4167
c. 0.2

1.31 c. Same range, no outliers.

1.33 b. Relatively mound-shaped, centred at 5.2.
c. Somewhat unusual.

1.35 a. Skewed right, with two outliers.
b. Dot plot is more informative; better display of the data shape with outliers shown.

1.37 c. Pareto chart seems more effective since it is very easy to compare the relative membership of the organized religions.

1.39 a. skewed   b. symmetric   c. symmetric
d. symmetric   e. skewed   f. skewed

1.41 a. continuous   b. continuous   c. discrete
d. discrete   e. discrete

1.43

7	8	9								
8	0	1	7							
9	0	1	2	4	4	5	6	6	6	8	8
10	1	7	9							
11	2									

1.45 a. skewed right

No Exercises for Section 5.2

Section 5.3 : Exercises

78   CHAPTER 4   Displaying and Summarizing Quantitative Data

EXERCISES

1. **Statistics in print.**   Find a histogram that shows the distribution of a variable in a newspaper or magazine article.
   a) Does the article identify the W's?
   b) Discuss whether the display is appropriate for the data.
   c) Discuss what the display reveals about the variable and its distribution.
   d) Does the article accurately describe and interpret the data? Explain.

2. **Not a histogram.**   Find a graph other than a histogram that shows the distribution of a quantitative variable in a newspaper or magazine article.
   a) Does the article identify the W's?
   b) Discuss whether the display is appropriate for the data.
   c) Discuss what the display reveals about the variable and its distribution.
   d) Does the article accurately describe and interpret the data? Explain.

3. **In the news.**   Find an article in a newspaper or a magazine that discusses an "average."
   a) Does the article discuss the W's for the data?
   b) What are the units of the variable?
   c) Is the average used the median or the mean? How can you tell?
   d) Is the choice of median or mean appropriate for the situation? Explain.

4. **In the news II.**   Find an article in a newspaper or a magazine that discusses a measure of spread.
   a) Does the article discuss the W's for the data?
   b) What are the units of the variable?
   c) Does the article use the range, IQR, or standard deviation?
   d) Is the choice of measure of spread appropriate for the situation? Explain.

5. **Thinking about shape.**   Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.
   a) The number of speeding tickets each student in the senior class of a college has ever had.
   b) Players' scores (number of strokes) at the U.S. Open golf tournament in a given year.
   c) Weights of female babies born in a particular hospital over the course of a year.
   d) The length of the average hair on the heads of students in a large class.

6. **More shapes.**   Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.
   a) Ages of people at a Little League game.
   b) Number of siblings of people in your class.
   c) Pulse rates of college-age males.
   d) Number of times each face of a die shows in 100 tosses.

7. **Sugar in cereals.**   The histogram displays the sugar content (as a percent of weight) of 49 brands of breakfast cereals.
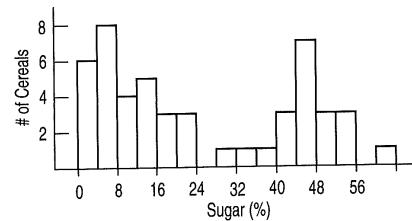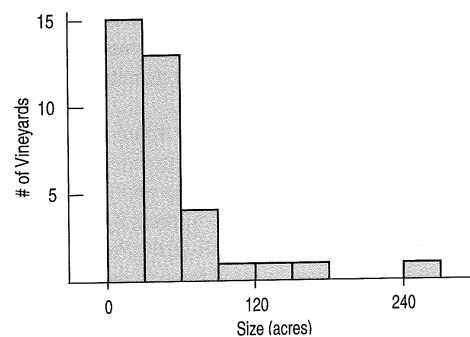

   a) Describe this distribution.
   b) What do you think might account for this shape?

8. **Singers.**   The display shows the heights of some of the singers in a chorus, collected so that the singers could be positioned on stage with shorter ones in front and taller ones in back.
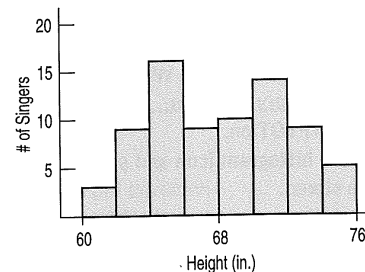
   a) Describe the distribution.
   b) Can you account for the features you see here?

9. **Vineyards.**   The histogram shows the sizes (in acres) of 36 vineyards in the Finger Lakes region of New York.

   a) Approximately what percentage of these vineyards are under 60 acres?
   b) Write a brief description of this distribution (shape, center, spread, unusual features).

5.3 Exercises

10. Run times. One of the authors collected the times (in minutes) it took him to run 4 miles on various courses during a 10-year period. Here is a histogram of the times.
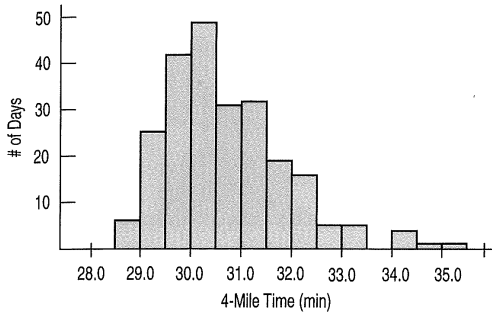
Describe the distribution and summarize the important features. What is it about running that might account for the shape you see?

11. Heart attack stays. The histogram shows the lengths of hospital stays (in days) for all the female patients admitted to hospitals in New York during one year with a primary diagnosis of acute myocardial infarction (heart attack).
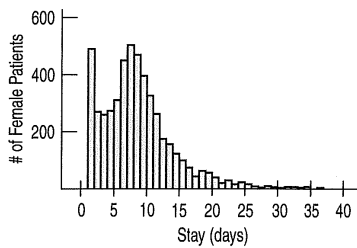
a) From the histogram, would you expect the mean or median to be larger? Explain.

b) Write a few sentences describing this distribution (shape, center, spread, unusual features).

12. E-mails. A university teacher saved every e-mail received from students in a large Introductory Statistics class during an entire term. He then counted, for each student who had sent him at least one e-mail, how many e-mails each student had sent.
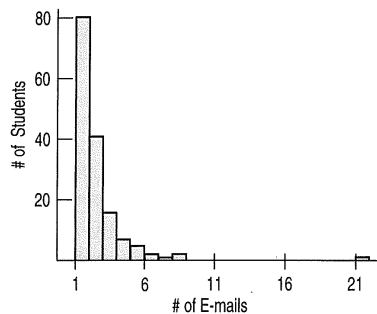
a) From the histogram, would you expect the mean or the median to be larger? Explain.

b) Write a few sentences describing this distribution (shape, center, spread, unusual features).

13. Super Bowl points. How many points do football teams score in the Super Bowl? Here are the total numbers of points scored by both teams in each of the first 41 Super Bowl games:

45, 47, 23, 30, 29, 27, 21, 31, 22, 38, 46, 37, 66, 50, 37, 47, 44, 47, 54, 56, 59, 52, 36, 65, 39, 61, 69, 43, 75, 44, 56, 55, 53, 39, 41, 37, 69, 61, 45, 31, 46

a) Find the median.
b) Find the quartiles.

14. Super Bowl wins. In the Super Bowl, by how many points does the winning team outscore the losers? Here are the winning margins for the first 41 Super Bowl games:

25, 19, 9, 16, 3, 21, 7, 17, 10, 4, 18, 17, 4, 12, 17, 5, 10, 29, 22, 36, 19, 32, 4, 45, 1, 13, 35, 17, 23, 10, 14, 7, 15, 7, 27, 3, 27, 3, 3, 11, 12

a) Find the median.
b) Find the quartiles.

15. Oil production. The histogram shows amount of oil produced (in gallons) from an acre of land in the United States from 36 different crops.
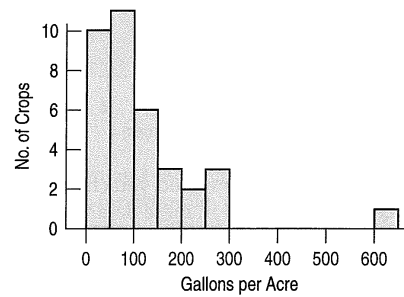
Which summary statistics would you choose to summarize the center and spread in these data? Why?

16. Paper consumption. The histogram shows the 2004 per capita consumption of paper for 195 countries around the world (in *kg per person per year*). (www.swivel.com)
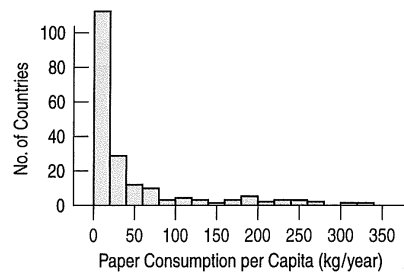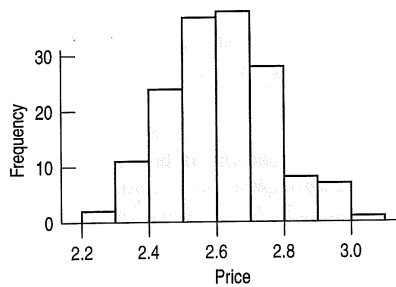

Which summary statistics would you choose to summarize the center and spread in these data? Why?

17. Pizza prices. The histogram shows the distribution of the prices of a small, plain pizza (in $) for 156 weeks in Dallas, Texas.
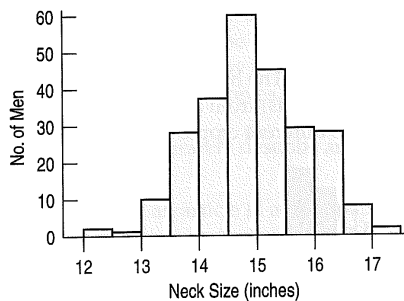
5.3 Exercises

Which summary statistics would you choose to summarize the center and spread in these data? Why?

18. Neck size. The histogram shows the neck sizes (in inches) of 250 men recruited for a health study in Utah.

Which summary statistics would you choose to summarize the center and spread in these data? Why?

19. Pizza prices again. Look again at the histogram of the pizza prices in Exercise 17.
a) Is the mean closer to $2.40, $2.60, or $2.80? Why?
b) Is the standard deviation closer to $0.15, $0.50, or $1.00? Explain.

20. Neck sizes again. Look again at the histogram of men's neck sizes in Exercise 18.
a) Is the mean closer to 14, 15, or 16 inches? Why?
b) Is the standard deviation closer to 1 inch, 3 inches, or 5 inches? Explain.

21. Movie lengths. The histogram shows the running times in minutes of 122 feature films released in 2005.
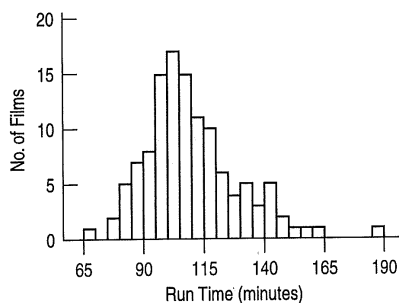
a) You plan to see a movie this weekend. Based on these movies, how long do you expect a typical movie to run?
b) Would you be surprised to find that your movie ran for 2 1/2 hours (150 minutes)?

c) Which would you expect to be higher, the mean or the median run time for all movies? Why?

22. Golf drives. The display shows the average drive distance (in yards) for 202 professional golfers on the men's PGA tour.
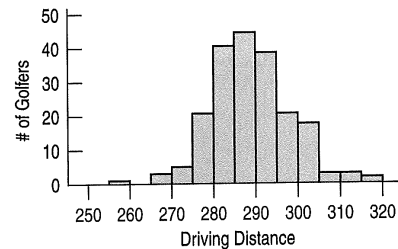
a) Describe this distribution.
b) Approximately what proportion of professional male golfers drive, on average, less than 280 yards?
c) Estimate the mean by examining the histogram.
d) Do you expect the mean to be smaller than, approximately equal to, or larger than the median? Why?

23. Summaries. Here are costs of 10 electric smoothtop ranges rated very good or excellent by *Consumer Reports* in August 2002:

$850 900 1400 1200 1050 1000 750 1250 1050 565

Find these statistics *by hand* (no calculator!):
a) mean
b) median and quartiles
c) range and IQR

24. More summaries. Here are the annual numbers of deaths from tornadoes in the United States from 1990 through 2000 (www.noaa.gov):

53 39 39 33 69 30 25 67 130 94 40

Find these statistics *by hand* (no calculator!):
a) mean
b) median and quartiles
c) range and IQR

25. Mistake. A clerk entering salary data into a company spreadsheet accidentally put an extra "0" in the boss's salary, listing it as $2,000,000 instead of $200,000. Explain how this error will affect these summary statistics for the company payroll:
a) measures of center: median and mean.
b) measures of spread: range, IQR, and standard deviation.

26. Cold weather. A meteorologist preparing a talk about global warming compiled a list of weekly low temperatures (in degrees Fahrenheit) he observed at his southern Florida home last year. The coldest temperature for any week was 36°F but he inadvertently recorded the Celsius value of 2°. Assuming that he correctly listed all the other temperatures, explain how this error will affect these summary statistics:
a) measures of center: mean and median.
b) measures of spread: range, IQR, and standard deviation.

    c) Probably unimodal and symmetric. Weights may be equally likely to be over or under the average.

    d) Probably bimodal. Men's and women's distributions may have different modes. It may also be skewed to the right, since it is possible to have very long hair, but hair length can't be negative.

7. a) Bimodal. Looks like two groups. Modes are near 6% and 46%. No real outliers.

    b) Looks like two groups of cereals, a low-sugar and a high-sugar group.

9. a) 78%

    b) Skewed to the right with at least one high outlier. Most of the vineyards are less than 90 acres with a few high ones. The mode is between 0 and 30 acres.

11. a) Because the distribution is skewed to the right, we expect the mean to be larger.

    b) Bimodal and skewed to the right. Center mode near 8 days. Another mode at 1 day (may represent patients who didn't survive). Most of the patients stay between 1 and 15 days. There are some extremely high values above 25 days.

13. a) 45 points   b) 37 points and 55 (or 55.5) points

15. The median and IQR because the distribution is strongly skewed.

17. The mean and standard deviation because the distribution is unimodal and symmetric.

19. a) The mean is closest to $2.60 because that's the balancing point of the histogram.

    b) The standard deviation is closest to $0.15 since that's a typical distance from the mean. There are no prices as far as $0.50 or $1.00.

21. a) About 100 minutes

    b) Yes, only 3 of these movies run that long.

    c) The mean would be higher. The distribution is skewed high.

23. a) $1001.50   b) 1025, 850, 1200   c) 835, 350

25. a) The median will probably be unaffected. The mean will be larger.

    b) The range and standard deviation will increase; the IQR will be unaffected.

27. The publication is using the median; the watchdog group is using the mean, pulled higher by the several very expensive movies in the long right tail.

29. a) The standard deviation will be larger for set 2, since the values are more spread out. SD(set 1) = 2.2, SD(set 2) = 3.2.

    b) The standard deviation will be larger for set 2, since 11 and 19 are farther from 15 than are 14 and 16. Other numbers are the same. SD(set 1) = 3.6, SD(set 2) = 4.5.

    c) The standard deviation will be the same for both sets, since the values in the second data set are just the values in the first data set + 80. The spread has not changed. SD(set 1) = 4.2. SD(set 2) = 4.2.

31. a) Mean $525, median $450

    b) 2 employees earn more than the mean.

    c) The median because of the outlier.

    d) The IQR will be least sensitive to the outlier of $1200, so it would be the best to report.

33. a)

Stem	Leaf
25	
25	
24	5 6
24	
23	6 8
23	2 3
22	6 7 7 7 8 9
22	1 2 3 4

22 | 1 = $2.21/gallon

    b) The distribution of gas prices is unimodal and skewed to the right (upward), centered around $2.27, with most stations

Section 5.3 Answers

CHAPTER 4

1. Answers will vary.

3. Answers will vary.

5. a) Unimodal (near 0) and skewed. Many seniors will have 0 or 1 speeding tickets. Some may have several, and a few may have more than that.

    b) Probably unimodal and slightly skewed to the right. It is easier to score 15 strokes over the mean than 15 strokes under the mean.

values. This in turn makes comparison of the data difficult. The only dimension that matters is bar height, so this is what should be emphasized. Figure 32 displays the same data in a two-dimensional bar graph. Which graphic is easier to read?

Figure 32   **Active Duty Personnel**
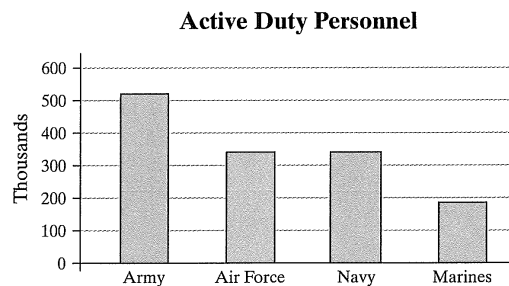
The material presented in this section is by no means all-inclusive. There are many ways to create graphs that mislead. Two popular texts written about ways that graphs mislead or deceive are *How to Lie with Statistics* (W. W. Norton & Company, Inc., 1982) by Darrell Huff and *The Visual Display of Quantitative Information* (Graphics Press, 2001) by Edward Tufte.

We conclude this section with some guidelines for constructing good graphics.

- Title and label the graphic axes clearly, providing explanations if needed. Include units of measurement and a data source when appropriate.
- Avoid distortion. Never lie about the data.
- Minimize the amount of white space in the graph. Use the available space to let the data stand out. If scales are truncated, be sure to clearly indicate this to the reader.
- Avoid clutter, such as excessive gridlines and unnecessary backgrounds or pictures. Don't distract the reader.
- Avoid three dimensions. Three-dimensional charts may look nice, but they distract the reader and often lead to misinterpretation of the graphic.
- Do not use more than one design in the same graphic. Sometimes graphs use a different design in one portion of the graph to draw attention to that area. Don't try to force the reader to any specific part of the graph. Let the data speak for themselves.
- Avoid relative graphs that are devoid of data or scales.

Section 5.4 Exercises

2.4 ASSESS YOUR UNDERSTANDING

Applying the Concepts

1. **Inauguration Cost** The following is a *USA Today*-type graph. Explain how it is misleading.
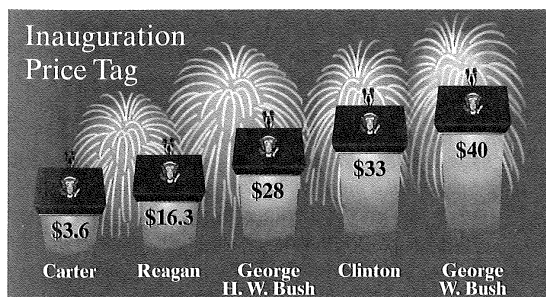
2. **Burning Calories** The following is a *USA Today*-type graph.

(a) Explain how it is misleading.
(b) What could be done to improve the graphic?

5.4 Exercises

116 Chapter 2 Organizing and Summarizing Data

3. Median Earnings The following graph shows the median earnings for females from 2002 to 2006 in constant 2006 dollars.
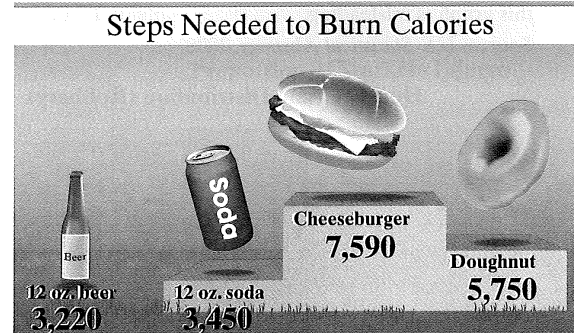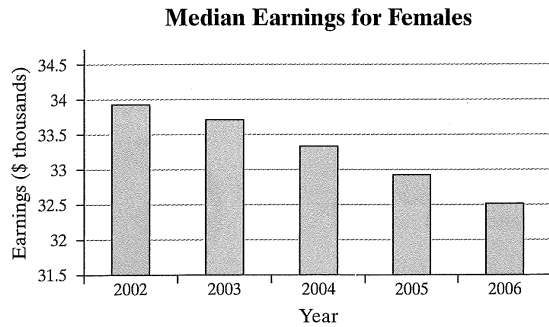*Source*: U.S. Census Bureau, Income, Poverty, and Health Insurance Coverage in the United States, 2006

Median Earnings for Females

(a) How is the bar graph misleading? What does the graph seem to convey?
(b) Redraw the graph so that it is not misleading. What does the new graph seem to convey?

4. Union Membership The following relative frequency histogram represents the proportion of employed people aged 25 to 64 years old who were members of a union.
Source: U.S. Bureau of Labor Statistics
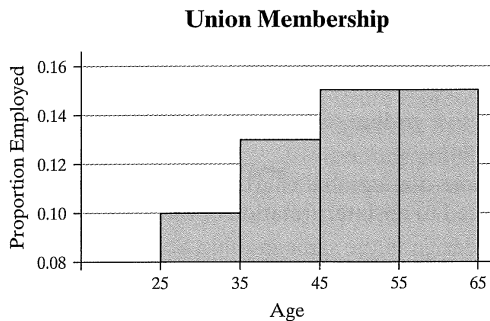
Union Membership

(a) Describe how this graph is misleading. What might a reader conclude from the graph?
(b) Redraw the histogram so that it is not misleading.

5. Robberies A newspaper article claimed that the afternoon hours were the worst in terms of robberies and provided the following graph in support of this claim. Explain how this graph is misleading.
Source: U.S. Statistical Abstract, 2008

Hourly Crime Distribution (Robbery)
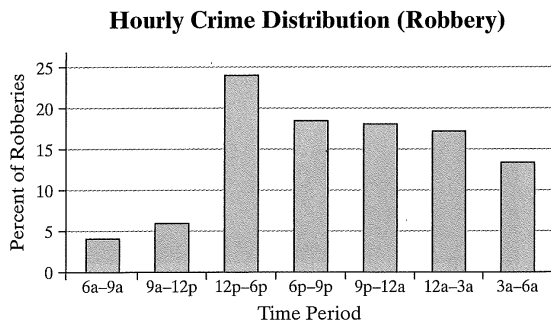
6. Car Accidents An article in a student newspaper claims that younger drivers are safer than older drivers and provides the following graph to support the claim. Explain how this graph is misleading.
Source: U.S. Statistical Abstract, 2008

Number of Motor Vehicle Accidents, 2005
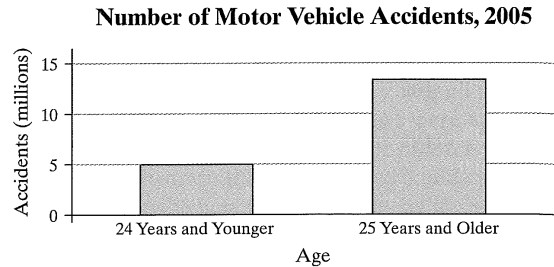
7. Health Insurance The following relative frequency histogram represents the proportion of people aged 25 to 64 years old not covered by any health insurance in 2006.
Source: U.S. Census Bureau

Proportion Not Covered by Health Insurance
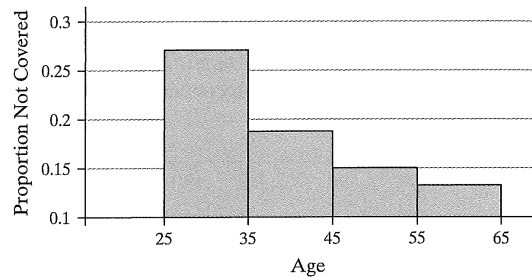
(a) Describe how this graph is misleading. What might a reader conclude from the graph?
(b) Redraw the histogram so that it is not misleading.

8. New Homes The following time-series plot shows the number of new homes built in the Midwest from 2000 to 2006.
Source: U.S. Statistical Abstract, 2008

New Homes in Midwest
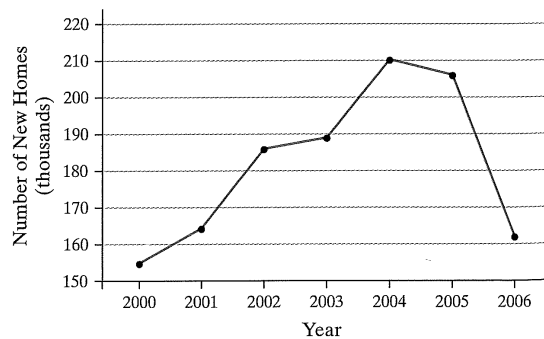
(a) Describe how this graph is misleading.
(b) What is the graph trying to convey?
(c) In January 2006, the National Association of Realtors reported, "*A lot of demand has been met over the last five years, and a modest rise in mortgage interest rates is causing some market cooling. Along with regulatory tightening on nontraditional mortgages, there will be fewer investors in the market this year.*" Does the graph support this view? Explain why or why not.

S.4 Exercises

9. **Median Income** The following time-series plot shows the median household income for the years 2001 to 2006 in constant 2006 dollars.

Source: U.S. Census Bureau

U.S. Median Household Income
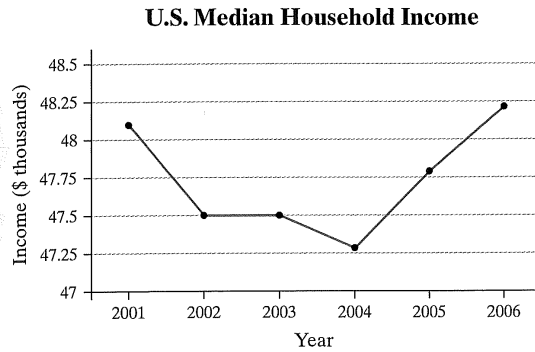
(a) Describe how the graph is misleading.

(b) What is the graph trying to convey?

(c) Redraw the graph so that Median Household Income appears to be relatively stable for the years shown.

10. **You Explain It! Oil Reserves** The U.S. Strategic Oil Reserve is a government-owned stockpile of crude oil. It was established after the oil embargo in the mid-1970s and is meant to serve as a national defense fuel reserve, as well as to offset reductions in commercial oil supplies that would threaten the U.S. economy.

Source: U.S. Energy Information Administration

**U.S. Strategic Oil Reserves
(millions of barrels)**

696.3

7.5

1977   2007
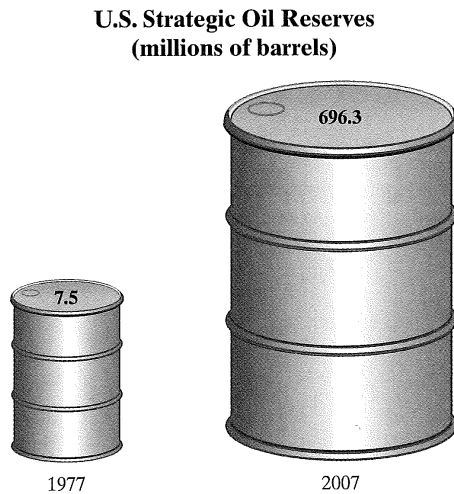
(a) How many times larger should the graphic for 2007 be than the 1977 graphic (to the nearest whole number)?

(b) The United States imported approximately 10.1 million barrels of oil per day in 2007. At that rate, assuming no change in U.S. oil production, how long would the U.S. strategic oil reserve last if no oil were imported?

11. **Cost of Kids** The following is a *USA Today*-type graph based on data from the Department of Agriculture. It represents the percentage of income a middle-income family will spend on their children.

Cost of Raising Kids

Housing	**33%**
Food	**18%**
Transportation	**15%**
Other	**11%**
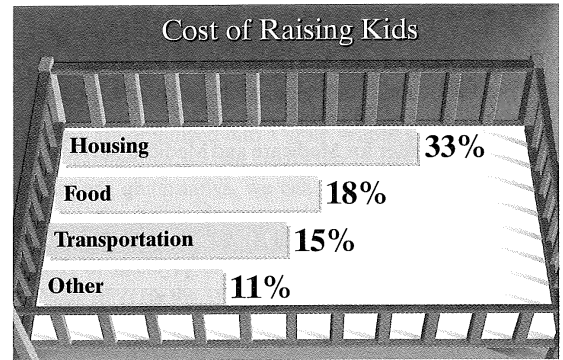
(a) How is the graphic misleading?

(b) What could be done to improve the graphic?

**12.** **Electricity** The following table gives the average per kilowatt-hour prices of electricity in the United States for the years 2001 to 2007.

Source: U.S. Energy Information Administration

Year	2001	2002	2003	2004	2005	2006	2007
Price per kWh (cents)	8.58	8.44	8.72	8.95	9.45	10.40	10.65

(a) Construct a misleading graph indicating that the price per kilowatt-hour has more than tripled since 2001.

(b) Construct a graph that is not misleading.

13. **ACT Composite** The following table gives the average ACT composite scores for the years 2003–2007.

Year	2003	2004	2005	2006	2007
Average ACT composite	20.8	20.9	20.9	21.1	21.2

(a) Construct a misleading time-series plot that indicates the average ACT composite score has risen sharply over the given time period.

(b) Construct a time-series plot that is not misleading.

(c) Which of the two graphs would you prefer if you were merely looking for trends in the data? Explain.

14. **Worker Injury** The safety manager at Klutz Enterprises provides the following graph to the plant manager and claims that the rate of worker injuries has been reduced by 67% over a 12-year period. Does the graph support his claim? Explain.

Proportion of Workers Injured
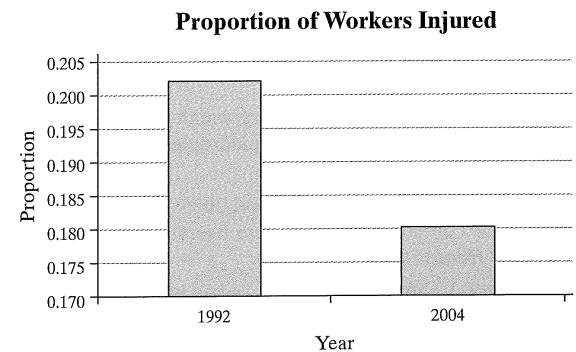
118   Chapter 2   Organizing and Summarizing Data

15. Health Care Expenditures The following data represent health care expenditures as a percentage of the U.S. gross domestic product (GDP) from 2001 to 2007. Gross domestic product is the total value of all goods and services created during the course of the year.

Source: Center for Medicare and Medicad Services, Office of the Actuary

Year	Health Care as a Percent of GDP
2001	14.5
2002	15.3
2003	15.8
2004	15.9
2005	16.0
2006	16.0
2007	16.2

(a) Construct a time-series plot that a politician would create to support the position that health care expenditures, as a percentage of GDP, are increasing and must be slowed.

(b) Construct a time-series plot that the health care industry would create to refute the opinion of the politician.

(c) Construct a time-series plot that is not misleading.

16. Motor Vehicle Death Rates The following data represent the number of motor vehicle deaths (within 30 days of accident) and the traffic death rates (number of deaths per 100,000 licensed drivers) from 2001 to 2005.

Year	Motor Vehicle Deaths (in thousands)	Traffic Death Rate (per 100,000 licensed drivers)
2001	42.2	22.1
2002	43.0	22.0
2003	42.9	21.9
2004	42.8	21.5
2005	43.4	21.7

Source: U.S. Statistical Abstract, 2008

(a) Construct a time-series graph to support the belief that the roads are becoming less safe.

(b) Construct a time-series graph to support the belief that the roads are becoming safer.

(c) Which graph do you feel better represents the situation?

17. Gas Hike The average per gallon price for regular unleaded gasoline in the United States rose from $1.46 in 2001 to $4.01 in 2008.

Source: U.S. Energy Information Administration

(a) Construct a graphic that is not misleading to depict this situation.

(b) Construct a misleading graphic that makes it appear the average price roughly quadrupled between 2001 and 2007.

18. Overweight Between 1980 and 2006, the number of adults in the United States who were overweight more than doubled from 15% to 34%.

Source: Centers for Disease Control and Prevention

(a) Construct a graphic that is not misleading to depict this situation.

(b) Construct a misleading graphic that makes it appear that the percent of overweight adults has more than quadrupled between 1980 and 2006.

19. Corn Production The following *USA Today*-type graphic illustrates U.S. corn production in billions of bushels for the years 1998 to 2007.
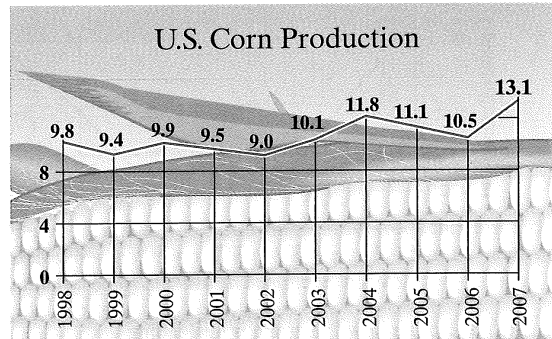
(a) What type of graph is being displayed?

(b) Describe some of the problems with this graphic.

(c) Construct a new graphic that is not misleading and makes the data stand out.

20. Putting It Together: College Costs The cover of the *Ithaca Times* from December 7, 2000 is shown.

The Rising Cost of College
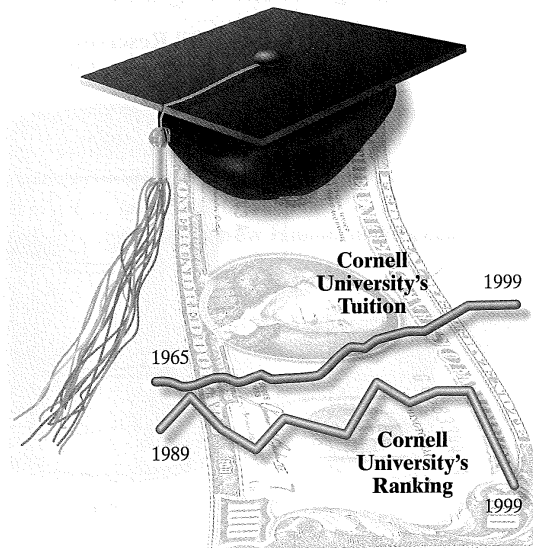
(a) Identify the two variables being graphed and describe them in terms of type and measurement level.

(b) What type of data collection method was likely used to create this graph?

(c) What type of graph is displayed?

(d) What message does the graph convey to you? How might this graph be misleading?

(e) Describe at least three things that are wrong with the graph.

ANS-14   Answers   2.3 Assess Your Understanding

Section 5.4:

Answers

2.4 Assess Your Understanding (page 115)

1. The lengths of the bars are not proportional. For example, the bar representing the cost of Clinton's inauguration should be slightly more than 9 times the one for Carter's cost and twice as long as the bar representing Reagan's cost.

3. **(a)** The vertical axis starts at 31.5 instead of 0. This tends to indicate that the median earnings for females decreased at a faster rate than they actually did.

(b) This graph indicates that the median earnings for females have decreased slightly over the given time period.

Median Earnings for Females
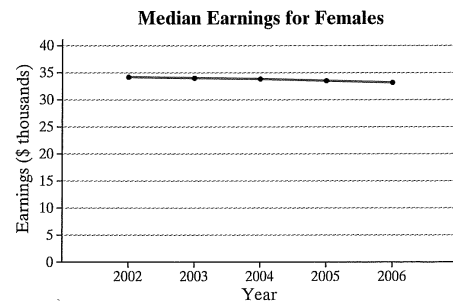
5. The bar for 12p–6p covers twice as many hours as the other bars. By combining two 3-hour periods, this bar looks larger compared to the others, making afternoon hours look more dangerous. When the bar is split into two periods, the graph may give a different impression.

7. **(a)** The vertical axis starts at 0.1 instead of 0. This might cause the reader to conclude, for example, that the proportion of people aged 25 to 34 who are not covered by health insurance is more than twice the proportion for those aged 45 to 54 years.

5.4 Answers

(b)

Proportion Not Covered by Health Insurance
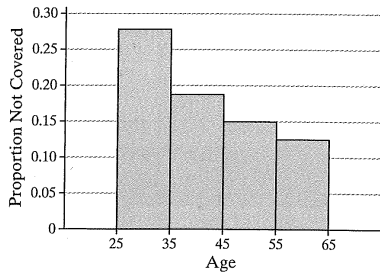
9. (a) The vertical axis starts at 47 without indicating a gap.
(b) It may convey that the median household income is increasing after a period of decline.

(c)

U.S. Median Household Income
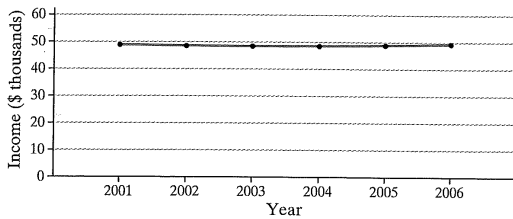
11. (a) The bar for housing should be a little more than twice the length of the bar for transportation, but it is not.
(b) Adjust the graph so that the lengths of the bars are proportional.

13. (a)

ACT Composite Score
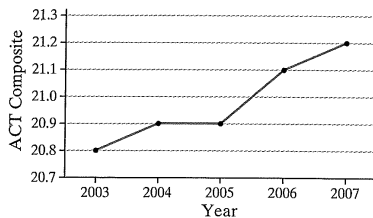
(b)

ACT Composite Score
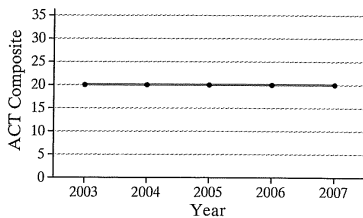
(c) Graph (a) is preferred because the trend can be seen.

15. (a) The politician's view:
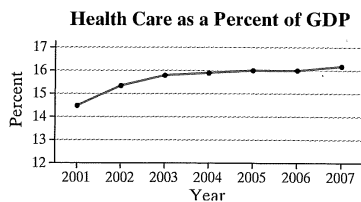
Health Care as a Percent of GDP

(b) The health care industry's view:

Health Care as a Percent of GDP
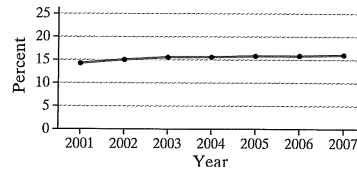
(c) A view that is not misleading:

Health Care as a Percent of GDP
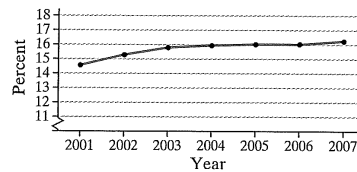
17. (a) Graphic that is not misleading:

Unleaded Gasoline Cost
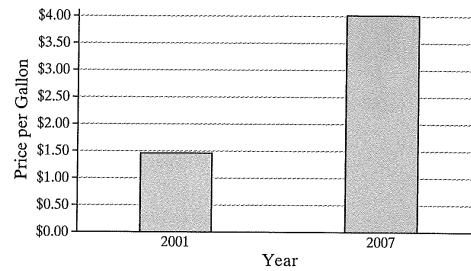
(b) Graphic that is misleading (graphics may vary):

Unleaded Gasoline Cost
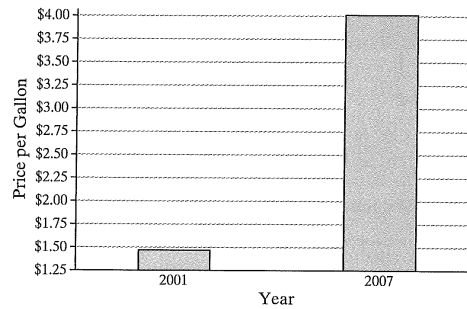
19. (a) Time series
(b) The graph is too cluttered; the axes are not labeled; the grid stands out more than the data.

(c) Graph that is not misleading:

U.S. Corn Production
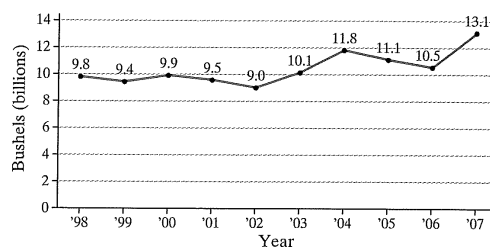
Summary

We conclude this section with the following chart, which addresses the circumstances under which each measure of central tendency should be used.

Measure of Central Tendency	Computation	Interpretation	When to Use
Mean	Population mean: $\mu = \dfrac{\Sigma x_i}{N}$ Sample mean: $\overline{x} = \dfrac{\Sigma x_i}{n}$	Center of gravity	When data are quantitative and the frequency distribution is roughly symmetric
Median	Arrange data in ascending order and divide the data set in half	Divides the bottom 50% of the data from the top 50%	When the data are quantitative and the frequency distribution is skewed left or skewed right

Section ~~2.1~~ 6.1 Exercises

3.1 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. What does it mean if a statistic is resistant? Why is the median resistant, but the mean is not?
2. In the 2000 census conducted by the U.S. Census Bureau, two average household incomes were reported: $41,349 and $55,263. One of these averages is the mean and the other is the median. Which is the mean? Support your answer.
3. The U.S. Department of Housing and Urban Development (HUD) uses the median to report the average price of a home in the United States. Why do you think HUD uses the median?
4. A histogram of a set of data indicates that the distribution of the data is skewed right. Which measure of central tendency will likely be larger, the mean or the median? Why?
5. If a data set contains 10,000 values arranged in increasing order, where is the median located?
6. *True or False*: A data set will always have exactly one mode.

Skill Building

In Problems 7–10, find the population mean or sample mean as indicated.

7. Sample: 20, 13, 4, 8, 10
8. Sample: 83, 65, 91, 87, 84
9. Population: 3, 6, 10, 12, 14
10. Population: 1, 19, 25, 15, 12, 16, 28, 13, 6
11. For Super Bowl XL, CBS television sold 65 ad slots for a total revenue of roughly $162.5 million. What was the mean price per ad slot?

12. The median for the given set of six ordered data values is 26.5. What is the missing value? 7 12 21 _____ 41 50
13. **Crash Test Results** The Insurance Institute for Highway Safety crashed the 2007 Audi A4 four times at 5 miles per hour. The costs of repair for each of the four crashes were

    $976, $2038, $918, $1899

    Compute the mean, median, and mode cost of repair.
14. **Cell Phone Use** The following data represent the monthly cell phone bill for my wife's phone for six randomly selected months.

    $35.34, $42.09, $39.43, $38.93, $43.39, $49.26

    Compute the mean, median, and mode phone bill.
15. **Concrete Mix** A certain type of concrete mix is designed to withstand 3,000 pounds per square inch (psi) of pressure. The strength of concrete is measured by pouring the mix into casting cylinders 6 inches in diameter and 12 inches tall. The concrete is allowed to set for 28 days. The concrete's strength is then measured. The following data represent the strength of nine randomly selected casts (in psi).

    3960, 4090, 3200, 3100, 2940, 3830, 4090, 4040, 3780

    Compute the mean, median, and mode strength of the concrete (in psi).
16. **Flight Time** The following data represent the flight time (in minutes) of a random sample of seven flights from Las Vegas, Nevada, to Newark, New Jersey, on Continental Airlines.

    282, 270, 260, 266, 257, 260, 267

    Compute the mean, median, and mode flight time.

6.1 Exercises

138   Chapter 3   Numerically Summarizing Data

17. For each of the three histograms shown, determine whether the mean is greater than, less than, or approximately equal to the median. Justify your answer.
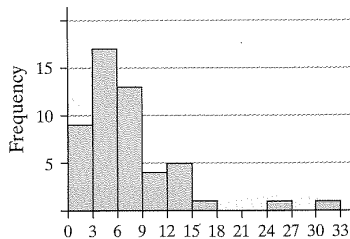
(a)
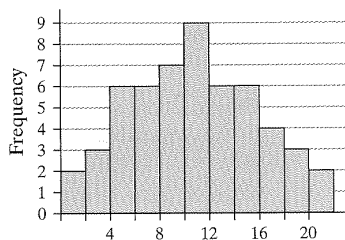
(b)
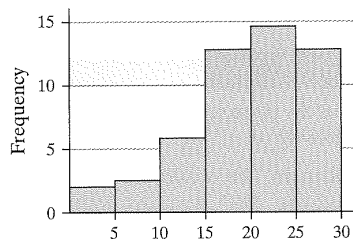
(c)

18. Match the histograms shown to the summary statistics:

	Mean	Median
I	42	42
II	31	36
III	31	26
IV	31	32
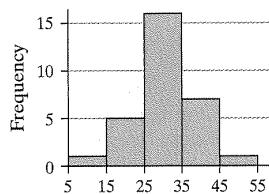
(a)
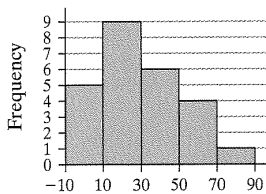
(b)
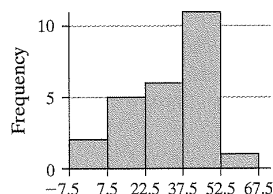
(c)
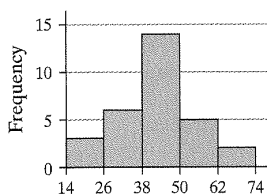
(d)

19. Mean versus Median Applet Load the mean versus median applet that is located on the CD that accompanies the text. Change the lower limit to 0 and the upper limit to 10 and click Update.

(a) Create a data set of at least eight observations such that the mean and median are roughly 2.

(b) Add a single observation near 9. How does this new value affect the mean? How does this new value affect the median?

(c) Change the upper limit to 25 and click Update. Remove the single value added from part (b). Add a single observation near 24. How does this new value affect the mean? the median?

(d) Refresh the page. Change the lower limit to 0 and the upper limit to 50 and click Update. Create a data set of at least eight observations such that the mean and median are roughly 40.

(e) Add a single observation near 35. How does this new value affect the mean? the median? Now "grab" this point with your mouse cursor and drag it toward 0. What happens to the value of the mean? What happens to the value of the median? Why?

20. Mean versus Median Applet Load the mean versus median applet that is located on the CD that accompanies the text. Change the lower limit to 0 and the upper limit to 10 and click Update.

(a) Create a data set of at least 10 observations such that the mean equals the median.

(b) Create a data set of at least 10 observations such that the mean is greater than the median.

(c) Create a data set of at least 10 observations such that the mean is less than the median.

(d) Comment on the shape of each distribution from parts (a)–(c).

(e) Can you create a distribution that is skewed left, but has a mean that is greater than the median?

Applying the Concepts

21. pH in Water The acidity or alkalinity of a solution is measured using pH. A pH less than 7 is acidic; a pH greater than 7 is alkaline. The following data represent the pH in samples of bottled water and tap water.

Tap	7.64	7.45	7.47	7.50	7.68	7.69
	7.45	7.10	7.56	7.47	7.52	7.47
Bottled	5.15	5.09	5.26	5.20	5.02	5.23
	5.28	5.26	5.13	5.26	5.21	5.24

Source: Emily McCarney, student at Joliet Junior College

(a) Compute the mean, median, and mode pH for each type of water. Comment on the differences between the two water types.

(b) Suppose the pH of 7.10 in tap water was incorrectly recorded as 1.70. How does this affect the mean? the median? What property of the median does this illustrate?

22. Reaction Time In an experiment conducted online at the University of Mississippi, study participants are asked to react to a stimulus. In one experiment, the participant must

6.1 Exercises

781	1,038	453	1,446	3,082
501	451	1,826	1,348	3,001
1,342	1,889	580	0	2,909
2,883	480	1,664	1,064	2,978
149	1,291	507	261	540
543	87	798	673	2,862
1,692	1,783	2,186	398	526
730	2,324	2,823	1,676	4,148

Source: Ashley Hudson, student at Joliet Junior College

Determine the shape of the distribution of new-car profit by drawing a frequency histogram. Compute the mean and median. Which measure of central tendency better describes the profit?

33. Political Views A sample of 30 registered voters was surveyed in which the respondents were asked, "Do you consider your political views to be conservative, moderate, or liberal?" The results of the survey are shown in the table.

Liberal	Conservative	Moderate
Moderate	Liberal	Moderate
Liberal	Moderate	Conservative
Moderate	Conservative	Moderate
Moderate	Moderate	Liberal
Liberal	Moderate	Liberal
Conservative	Moderate	Moderate
Liberal	Conservative	Liberal
Liberal	Conservative	Liberal
Conservative	Moderate	Conservative

Source: Based on data from the General Social Survey

(a) Determine the mode political view.
(b) Do you think it would be a good idea to rotate the choices conservative, moderate, or liberal in the question? Why?

34. Hospital Admissions The following data represent the diagnosis of a random sample of 20 patients admitted to a hospital.

Cancer	Motor vehicle accident	Congestive heart failure
Gunshot wound	Fall	Gunshot wound
Gunshot wound	Motor vehicle accident	Gunshot wound
Assault	Motor vehicle accident	Gunshot wound
Motor vehicle accident	Motor vehicle accident	Gunshot wound
Motor vehicle accident	Gunshot wound	Motor vehicle accident
Fall	Gunshot wound	

Source: Tamela Ohm, student at Joliet Junior College

Determine the mode diagnosis.

35. Resistance and Sample Size Each of the following three data sets represents the IQ scores of a random sample of adults. IQ scores are known to have a mean and median

of 100. For each data set, compute the mean and median. For each data set recalculate the mean and median, assuming that the individual whose IQ is 106 is accidentally recorded as 160. For each sample size, state what happens to the mean and the median? Comment on the role the number of observations plays in resistance.

Sample of Size 5				
106	92	98	103	100

Sample of Size 12					
106	92	98	103	100	102
98	124	83	70	108	121

Sample of Size 30					
106	92	98	103	100	102
98	124	83	70	108	121
102	87	121	107	97	114
140	93	130	72	81	90
103	97	89	98	88	103

36. Mr. Zuro finds the mean height of all 14 students in his statistics class to be 68.0 inches. Just as Mr. Zuro finishes explaining how to get the mean, Danielle walks in late. Danielle is 65 inches tall. What is the mean height of the 15 students in the class?

37. A researcher with the Department of Energy wants to determine the mean natural gas bill of households throughout the United States. He knows the mean natural gas bill of households for each state, so he adds together these 50 values and divides by 50 to arrive at his estimate. Is this a valid approach? Why or why not?

38. Net Worth According to the *Statistical Abstract of the United States*, the mean net worth of all households in the United States in 2004 was $448,200, while the median net worth was $93,100.
(a) Which measure do you believe better describes the typical U.S. household's net worth? Support your opinion.
(b) What shape would you expect the distribution of net worth to have? Why?
(c) What do you think causes the disparity in the two measures of central tendency?

39. You are negotiating a contract for the Players Association of the NBA. Which measure of central tendency will you use to support your claim that the average player's salary needs to be increased? Why? As the chief negotiator for the owners, which measure would you use to refute the claim made by the Players Association?

40. In January 2008, the mean amount of money lost per visitor to a local riverboat casino was $135. Do you think the median was more than, less than, or equal to this amount? Why?

41. Missing Exam Grade A professor has recorded exam grades for 20 students in his class, but one of the grades is no longer readable. If the mean score on the exam was 82 and the mean of the 19 readable scores is 84, what is the value of the unreadable score?

42. For each of the following situations, determine which measure of central tendency is most appropriate and justify your reasoning.

6.1 Exercises

142   Chapter 3   Numerically Summarizing Data

(a) Average price of a home sold in Pittsburgh, Pennsylvania in 2009
(b) Most popular major for students enrolled in a statistics course
(c) Average test score when the scores are distributed symmetrically
(d) Average test score when the scores are skewed right
(e) Average income of a player in the National Football League
(f) Most requested song at a radio station

43. Linear Transformations Benjamin owns a small Internet business. Besides himself, he employs nine other people. The salaries earned by the employees are given next in thousands of dollars (Benjamin's salary is the largest, of course):

$$30, 30, 45, 50, 50, 50, 55, 55, 60, 75$$

(a) Determine the mean, median, and mode for salary.
(b) Business has been good! As a result, Benjamin has a total of $25,000 in bonus pay to distribute to his employees. One option for distributing bonuses is to give each employee (including himself) $2,500. Add the bonuses under this plan to the original salaries to create a new data set. Recalculate the mean, median, and mode. How do they compare to the originals?
(c) As a second option, Benjamin can give each employee a bonus of 5% of his or her original salary. Add the bonuses under this second plan to the original salaries to create a new data set. Recalculate the mean, median, and mode. How do they compare to the originals?
(d) As a third option, Benjamin decides not to give his employees a bonus at all. Instead, he keeps the $25,000 for himself. Use this plan to create a new data set. Recalculate the mean, median, and mode. How do they compare to the originals?

44. Linear Transformations Use the five test scores of 65, 70, 71, 75, and 95 to answer the following questions:
(a) Find the sample mean.
(b) Find the median.
(c) Which measure of central tendency best describes the typical test score?

(d) Suppose the professor decides to curve the exam by adding 4 points to each test score. Compute the sample mean based on the adjusted scores.
(e) Compare the unadjusted test score mean with the curved test score mean. What effect did adding 4 to each score have on the mean?

45. Trimmed Mean Another measure of central tendency is the trimmed mean. It is computed by determining the mean of a data set after deleting the smallest and largest observed values. Compute the trimmed mean for the data in Problem 29. Is the trimmed mean resistant? Explain.

46. Midrange The midrange is also a measure of central tendency. It is computed by adding the smallest and largest observed values of a data set and dividing the result by 2; that is,

$$\text{Midrange} = \frac{\text{largest data value} + \text{smallest data value}}{2}$$

Compute the midrange for the data in Problem 29. Is the midrange resistant? Explain.

47. Putting It Together: Shape, Mean and Median As part of a semester project in a statistics course, Carlos surveyed a sample of 40 high school students and asked, "How many days in the past week have you consumed an alcoholic beverage?" The results of the survey are shown next.

0	0	1	4	1	1	1	5	1	3
0	1	0	1	0	4	0	1	0	1
0	0	0	0	2	0	0	0	0	0
1	0	2	0	0	0	1	2	1	1
2	0	1	0	1	3	1	1	0	3

(a) Is this data discrete or continuous?
(b) Draw a histogram of the data and describe its shape.
(c) Based on the shape of the histogram, do you expect the mean to be more than, equal to, or less than the median?
(d) Compute the mean and the median. What does this tell you?
(e) Determine the mode.
(f) Do you believe that Carlos' survey suffers from sampling bias? Why?

Section 6.1 Answers

CHAPTER 3 Numerically Summarizing Data

3.1 Assess Your Understanding (page 137)

1. A statistic is resistant if it is not sensitive to extreme values. The median is resistant because it is a positional measure of central tendency, and increasing the largest value or decreasing the smallest value does not affect the position of the center. The mean is not resistant because it is a function of the sum of the values of data. Changing the magnitude of one value changes the sum of the values.

3. HUD uses the median because the data are skewed. Explanations will vary.

5. The median is between the 5000th and the 5001st ordered values.

7. $\bar{x} = 11$

9. $\mu = 9$

11. Mean price per ad slot was $2.5 million.

13. Mean cost is $1,457.75; median cost is $1,437.50; there is no mode cost.

15. The mean, median, and mode strengths are 3,670, 3,830, and 4,090 pounds per square inch, respectively.

17. (a) mean > median
    (b) mean = median
    (c) mean < median
    Justification will vary.

21. (a) Tap: $\bar{x} = 7.50$; $M = 7.485$; mode = 7.47. Bottled: $\bar{x} = 5.194$; $M = 5.22$; mode = 5.26
    (b) $\bar{x} = 7.05$; $M = 7.485$; the median is resistant. *does not depend on outliers*

23. (a) The mean pulse rate is 72.2 beats per minute.
    (b) Samples and sample means will vary.
    (c) Answers will vary.

25. (a) 1,813,654.5 thousand metric tons
    (b) Per capita is better because it adjusts for CO_2 emissions for population. After all, countries with more people will, in general, have higher CO_2 emissions.
    (c) Mean = 2.814 thousand metric tons; median = 2.67 thousand metric tons; mean

27. The distribution is symmetric. The mean is the better measure of central tendency.

29. $\bar{x} = 0.875$ gram; $M = 0.875$ gram. The distribution is symmetric, so the mean is the better measure of central tendency.

Weight of Plain M&Ms
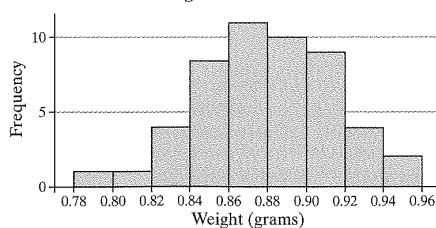
31. The distribution is skewed left; $\bar{x} = 22$ hours; $M = 25$ hours. The median is the better measure of central tendency.

Hours Worked per Week
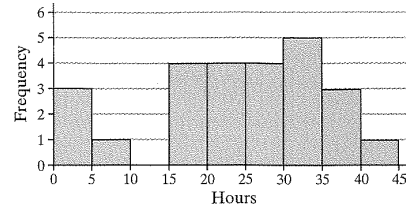
33. (a) Moderate   (b) Yes, to avoid response bias

35. Sample of size 5: All data recorded correctly: $\bar{x} = 99.8$; $M = 100$; 106 recorded as 160: $\bar{x} = 110.6$; $M = 100$

Sample of size 12: All data recorded correctly: $\bar{x} = 100.4$; $M = 101$; 106 recorded as 160: $\bar{x} = 104.9$; $M = 101$

Sample of size 30: All data recorded correctly: $\bar{x} = 100.6$; $M = 99$; 106 recorded as 160: $\bar{x} = 102.4$; $M = 99$

For each sample size, the mean becomes larger, but the median remains constant. As the sample size increases, the affect of the misrecorded data on the mean decreases.

37. No. Each state has a different population size. This must be taken into account.

39. The salary distribution is skewed right, so the players' negotiator would want to use the median salary; the owners' negotiator would use the mean salary to refute the players' claim.

41. The unreadable score is 44.

43. (a) Mean = $50,000; median = $50,000; mode = $50,000
    (b) New data set: 32.5, 32.5, 47.5, 52.5, 52.5, 52.5, 57.5, 57.5, 62.5, 77.5; mean = $52,500; median = $52,500; mode = $52,500. All three measures increased by $2,500.
    (c) New data set: 31.5, 31.5, 47.25, 52.5, 52.5, 52.5, 57.75, 57.75, 63, 78.75; mean = $52,500; median = $52,500; mode = $52,500. All three measures increased by 5%.
    (d) New data set: 30, 30, 45, 50, 50, 50, 55, 55, 60, 100; mean = $52,500; median = $50,000; mode = $50,000. The mean increased by $2,500, but the median and the mode remained at $50,000.

45. The trimmed mean is 0.875. Explanations will vary.

47. (a) Discrete
    (b)

Number of Days High School Students Consume Alcohol Each Week
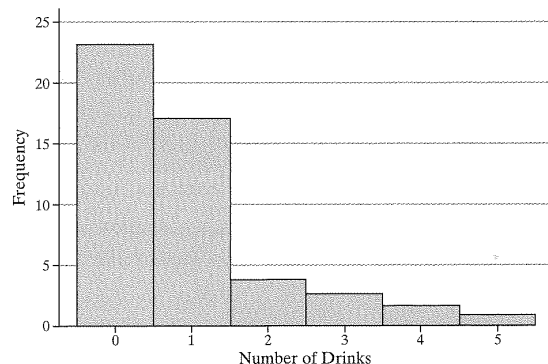
    (c) Since the data are skewed right, we would expect the mean to be greater than the median.
    (d) Mean: 0.94; median: 1; the mean can be less than the median in skewed-right data. Therefore, using the rule *mean greater than median implies the data are skewed right* does not always work.
    (e) 0
    (f) Yes. It is difficult to get truthful responses to this type of question. Carlos would need to ensure that the identity of the respondents is anonymous.

Section 6.2 Exercises

Section 6.4: Measures of Centre and of Variability

Exercises

1. The top ten movies (Skyfall, The Hobbit, etc.) and their profits (in millions of dollars) from last weekend are reported in Monday Magazine. Calculate the mean and median for this data.

   profits: 1.4, 4.1, 1.2, 1.3, 5.8, 5.0, 2.6, 1.8, 2.9, 5.9, 2.5, 5.3

2. Calculate the mean and median for the data set: 35, 47, 29, 42, 38, 39, 42.

3. Pat finds the mean height of all twelve students in her physics class to be 68.0 inches. Just as she's finished that calculation, one more student walks in late. If that student is 63.0 inches tall, what is the mean height of all thirteen students?

4. The Victoria Real Estate Board claims that in October of 2012, the average cost of a single-family home in Greater Victoria was $592,000, while the median was $527,000. Why is the mean greater than the median for housing prices? Explain.

5. Tom is running a small business with five employees, including himself. The salaries of the five people (in thousands of dollars) are 30, 45, 50, 55, and 75, with Tom making the highest salary.

   a) calculate the mean and median of these salaries
   b) if Tom gives everyone a $2000 bonus, what happens to the mean and median?
   c) if Tom gives everyone a 5% raise, what happens to the mean and median?
   d) if Tom decides to keep everyone else's salary the same, but raise his own salary by $10,000, what happens to the mean and the median?

6. Consider the following histogram. Is the standard deviation equal to

   a) 0.5
   b) 2
   c) 15
   d) 20

Random Numbers from the Standard Normal Distribution
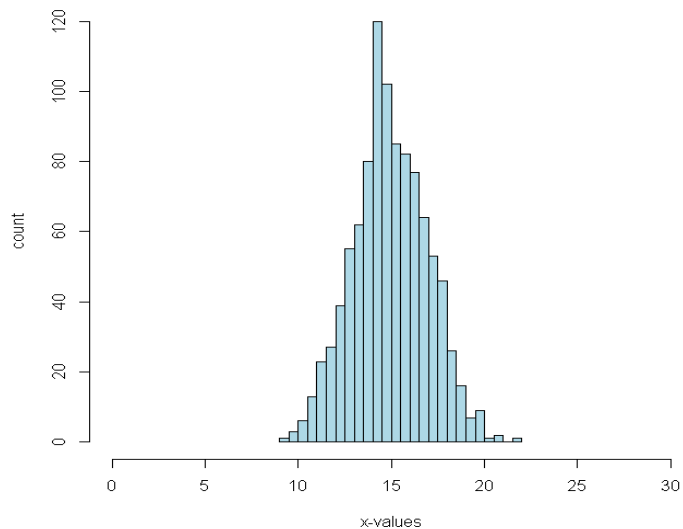
6.2 Exercises

7. Consider the following data set: 7, 7, 7, 7, 7, and 7. What is the mean and the median? What is the range? Without calculating it, what would be the standard deviation?

8. Consider the following histogram. Is the standard deviation equal to

a) 1
b) 2
c) 5
d) 10
e) 15
f) 20

Random Numbers from the Binomial Distribution
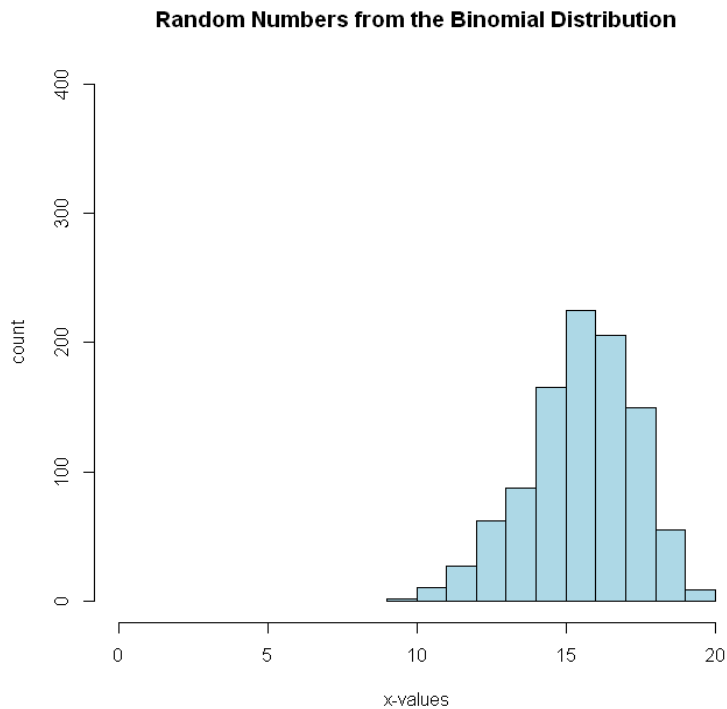
9. Pat, when entering quiz scores into her spreadsheet, accidentally put an extra zero on the end of one student's score (making it 380/40 instead of 38/40), and then calculated the mean, median, range, and standard deviation for the section. She then noticed her mistake and recalculated all of the quantities. For the following quantities, state whether the corrected value will be higher, lower, or the same as the value calculated with the incorrect quiz score:

a) mean
b) median
c) range
d) standard deviation

10. Consider the following sets of data. Without calculating any values, state which set will have the higher standard deviation (or will they be the same?).

a) Set 1: 2, 3, 9, 16, 17   Set 2: 2, 8, 9, 10, 17
b) Set 1: 2, 3, 9, 16, 17   Set 2: 3, 4, 10, 17, 18

Sections 6.2 Answers:

Section 6.4: Statistical Quantities

Solutions

1. The mean is 3.31667, or just 3.3. There are twelve points, so the median is the 12/2+1/2=6.5th point, which means the average of the 6th and 7th points. Therefore, the median is (2.6+2.9)/2 = 2.75.

2. The mean is 38.8571. (You can round to 38.9 if you like.)

3. To find the mean, we want the sum of all of the heights divided by the total number of students. Since the average of the twelve students is 68.0 inches, the total of all of those heights is just 68.0 times 12, which is 816.0 inches. Adding the height of the thirteenth student brings the total to 879.0 inches, then dividing by 13 gives a mean of 67.6 inches.

4. The histogram of Victoria housing prices will not be symmetrical: there is a lower limit for the price of single-family homes, while there can be house prices in the millions of dollars. Just a few very expensive homes will bring up the mean but not affect the median in any way, which is why the mean is greater than the median.

5. The means and medians are:

   a) mean = $51,000 and median = $50,000
   b) the mean and median will each increase by $2000: mean is now $53,000 and the median $52,000
   c) the mean and median will both increase by a factor of 1.05 (they are multiplied by 1.05): mean is now $53,550 and median is $52,500
   d) the mean will become $53,000 but the median will stay the same

6. Looking at the histogram, you can estimate the standard deviation by picking a "width" about the mean/average that most of the data points fall within. From this histogram, the standard deviation is about half of 5, since most of the data falls between approximately 12.5 and 17.5 (ish). And the closest value given that matches that is (b) 2.

7. The mean and median are both 7. The range is 0. The standard deviation is also 0, since all points lie exactly on the mean and $(x-\overline{x})$ is zero for each point.

8. Using the same reasoning as for question 6, most of the data seems to fall between 12.5 and 17.5, so the standard deviation is around 2.5 (ish). So the closest option given is (b) again.

9. New values:

   a) The corrected mean will be lower, since one value was lowered.

6.2 Answers

b) The median will remain unchanged (assuming that the 38/40 was in the upper half of the scores to begin with, so changing it to 380 and back won't affect that)

c) The corrected range will be lower, since the highest point has changed.

d) The standard deviation will be lower, since the corrected point's distance from the mean is lower than the uncorrected value.

10. a) Set 1's values are farther from the mean on average than Set 2's data points. So Set 1 will have a higher standard deviation.

b) Set 2's data points are just Set 1's points moved up by 1 unit. So each point's distance from the mean will be the same as Set 1, and the standard deviations will be the same also.

Section 6.3 : Exercises

116   Chapter 3 Data Description

26. Dog Reaction Times Find the variance and standard deviation for the two distributions in Exercise 8 in Section 2–2 and Exercise 18 in Section 2–2. Compare the variation of the data sets. Decide if one data set is more variable than the other.

27. Photocopier Service Calls This frequency distribution represents the data obtained from a sample of photocopier service technicians. The values are the days between service calls on 80 photocopy machines.

Days between calls	Frequency
25.5–28.5	5
28.5–31.5	9
31.5–34.5	32
34.5–37.5	20
37.5–40.5	12
40.5–43.5	2

28. Exam Scores The average score of the students in one calculus class is 110, with a standard deviation of 5; the average score of students in a statistics class is 106, with a standard deviation of 4. Which class is more variable in terms of scores?

29. Suspension Bridges The data show the lengths (in metres) of suspension bridges in the eastern part of North America and western part of North America. Compare the variability of the two samples.
East: 1298, 1067, 375, 655, 610, 533
West: 1250, 853, 704, 472, 457, 368
Source: World Almanac and Book of Facts.

30. Exam Scores The average score on an English final examination was 85, with a standard deviation of 5; the average score on a history final exam was 110, with a standard deviation of 8. Which class was more variable?

31. Accountants' Ages The average age of the accountants at Three Rivers Corp. is 26 years, with a standard deviation of 6 years; the average salary of the accountants is $31,000, with a standard deviation of $4000. Compare the variations of age and income.

32. Using Chebyshev's theorem, solve these problems for a distribution with a mean of 80 and a standard deviation of 10.
   a. At least what percentage of values will fall between 60 and 100?
   b. At least what percentage of values will fall between 65 and 95?

33. The mean of a distribution is 20 and the standard deviation is 2. Use Chebyshev's theorem.
   a. At least what percentage of the values will fall between 10 and 30?
   b. At least what percentage of the values will fall between 12 and 28?

34. In a distribution of 200 values, the mean is 50 and the standard deviation is 5. Use Chebyshev's theorem.
   a. At least how many values will fall between 30 and 70?
   b. At most how many values will be less than 40 or more than 60?

35. Fast-Food Industry Wages A sample of hourly wages of employees in the fast-food industry has a mean of $8.26 and a standard deviation of $0.33. Using Chebyshev's theorem, find the range in which at least 75% of the data values will fall.

36. Time Spent Online Adult Canadians spend an average of 2.7 hours per day online. Assuming a standard deviation of 30 minutes, find the range in which at least 88.89% of adult Canadian users spend online. Use Chebyshev's theorem.
Source: Ipsos, *News and Polls,* "Canadian Teenagers Are Leading the Online Revolution? Maybe Not...," February 27, 2008.

37. Cereal Potassium per Serving A survey of a number of the leading brands of cereal shows that the mean content of potassium per serving is 95 milligrams, and the standard deviation is 2 milligrams. Find the range in which at least 88.89% of the data will fall. Use Chebyshev's theorem.

38. Solid Waste Production The average college student produces 290 kilograms of solid waste each year, including 500 disposable cups and 140 kilograms of paper. If the standard deviation is approximately 36 kilograms, within what weight limits will at least 75% of all students' garbage lie?
Source: Environmental Sustainability Committee, www.esc.mtu.edu.

39. Trials to Learn a Maze The average of the number of trials it took a sample of mice to learn to traverse a maze was 12. The standard deviation was 3. Using Chebyshev's theorem, find the minimum percentage of data values that will fall in the range of 4 to 20 trials.

40. Farm Size The average farm in Canada in 2005 contained 295 hectares. Assume a standard deviation of 16 hectares. Use Chebyshev's theorem to find the minimum percentage of farms that fell in the range of 255 to 335 hectares.
Source: Agriculture and Agri-Food Canada, *Special Features: Census of Agriculture Summary.*

41. Fresh Food Consumption The average yearly per capita consumption of fresh fruit in Canada is 68.8 kilograms. Suppose that the distribution of fruit amounts consumed is bell-shaped with a standard deviation equal to 3.1 kilograms. What percentage of Canadians would you expect to consume more than 75 kilograms of fresh fruit per year?
Source: Statistics Canada.

42. Faculty Work Hours The average full-time faculty member in a post-secondary degree-granting institution works an average of 53 hours per week.
   a. If we assume the standard deviation is 2.8 hours, what percentage of faculty members work more than 58.6 hours a week?
   b. If we assume a bell-shaped distribution, what percentage of faculty members work more than 58.6 hours a week?
Source: National Center for Education Statistics.

6.3 Exercises

Extending the Concepts »

43. Serum Cholesterol Levels For this data set, find the mean and standard deviation of the variable. The data represent the serum cholesterol levels of 30 individuals. Count the number of data values that fall within 2 standard deviations of the mean. Compare this with the number obtained from Chebyshev's theorem. Comment on the answer.

211	240	255	219	204
200	212	193	187	205
256	203	210	221	249
231	212	236	204	187
201	247	206	187	200
237	227	221	192	196

44. Ages of Consumers For this data set, find the mean and standard deviation of the variable. The data represent the ages of 30 customers who ordered a product advertised on television. Count the number of data values that fall within 2 standard deviations of the mean. Compare this with the number obtained from Chebyshev's theorem. Comment on the answer.

42	44	62	35	20
30	56	20	23	41
55	22	31	27	66
21	18	24	42	25
32	50	31	26	36
39	40	18	36	22

45. Using Chebyshev's theorem, complete the table to find the minimum percentage of data values that fall within k standard deviations of the mean.

k	1.5	2	2.5	3	3.5
Percentage					

46. Use this data set: 10, 20, 30, 40, 50.
  a. Find the standard deviation.
  b. Add 5 to each value, and then find the standard deviation.
  c. Subtract 5 from each value and find the standard deviation.

d. Multiply each value by 5 and find the standard deviation.
e. Divide each value by 5 and find the standard deviation.
f. Generalize the results of parts b through e.
g. Compare these results with those in Exercise 38.

47. The mean deviation is found by using this formula:

$$\text{Mean deviation} = \frac{\Sigma |X - \overline{X}|}{n}$$

where
  X = value
  $\overline{X}$ = mean
  n = number of values
  $||$ = absolute value

Find the mean deviation for these data.

5, 9, 10, 11, 11, 12, 15, 18, 20, 22

48. A measure to determine the skewness of a distribution is called the *Pearson coefficient of skewness*. The formula is

$$\text{Skewness} = \frac{3(\overline{X} - \text{MD})}{s}$$

The values of the coefficient usually range from -3 to $+3$. When the distribution is symmetric, the coefficient is zero; when the distribution is positively skewed, it is positive; and when the distribution is negatively skewed, it is negative.

  Using the formula, find the coefficient of skewness for each distribution, and describe the shape of the distribution.

  a. Mean = 10, median = 8, standard deviation = 3.
  b. Mean = 42, median = 45, standard deviation = 4.
  c. Mean = 18.6, median = 18.6, standard deviation = 1.5.
  d. Mean = 98, median = 97.6, standard deviation = 4.

49. All values of a data set must be within $s\sqrt{n-1}$ of the mean. If a person collected 25 data values that had a mean of 50 and a standard deviation of 3 and you saw that one data value was 67, what would you conclude?

6.3
Section 2.3: Answers

17. $R = 11{,}263$; $7{,}436{,}475.0$; 2727.0

19. 133.6; 11.6 **21.** 45.93; 6.78

23. 211.2; 14.5 **25.** 211.2; 14.5;

No, the variability of the lifetimes of the batteries is quite large.

27. 11.7; 3.4

29. For West, CVar $= 46.4\%$. For East, CVar $= 48.3\%$. The data for East are more variable.

31. 23.1%; 12.9%.

The age is more variable.

33. *a.* 96% *b.* 93.75%

35. $\$7.60$–$\$8.92$ **37.** 89–101

39. 86% **41.** 2.5%

43. $n = 30$ $\overline{X} = 214.97$ $s = 20.76$.   At least 75% of the data values will fall between $\overline{X} \pm 2s$.
$\overline{X} - 2(20.76) = 214.97 - 41.52 = 173.45$ and
$\overline{X} + 2(20.76) = 214.97 + 41.52 = 256.49$

In this case all 30 values fall within this range; hence Chebyshev's theorem is correct for this example.

45. 56%; 75%; 84%; 88.89%; 92%

47. 4.36

49. It must be an incorrect data value, since it is beyond the range using the formula $s\sqrt{n-1}$.

Section 6.4 Exercises

Exercises 3–3

1. What is a z score?

2. Define *percentile rank.*

3. What is the difference between a percentage and a percentile?

4. Define *quartile.*

5. What is the relationship between quartiles and percentiles?

6. What is a decile?

7. How are deciles related to percentiles?

8. To which percentile, quartile, and decile does the median correspond?

9. **Vacation Days** If the average number of vacation days for a selection of various countries has a mean of 29.4 days and a standard deviation of 8.6, find the z scores for the average number of vacation days in each of these countries.

Canada	26 days
Italy	42 days
United States	13 days

Source: Infoplease: www.infoplease.com.

10. **Reaction Time of Sprinters** The mean reaction time to the starting pistol for world-class sprinters is 153 milliseconds (ms) with a standard deviation of 28 ms. Find the corresponding z score for each sprinter's reaction time (in ms).

 a. 195   *d.* 241.2
 b. 90   *e.* 88.6
 c. 139

Source: Kevin Duffy's Home Page, "Reaction Times and Sprint False Starts."

11. **Exam Scores** A final examination for a psychology course has a mean of 84 and a standard deviation of 4. Find the corresponding z score for each raw score.

 a. 87   *d.* 76
 b. 79   *e.* 82
 c. 93

12. **Temperature of the Human Body** The healthy human body has a mean temperature of 36.8°C, with a standard deviation of 0.7°C. Find the corresponding z score for the following body temperatures (°C).

 a. 37.5
 b. 36.1
 c. 34.77
 d. 37.85
 e. 38.41

Source: NationMaster.com, Normal Human Body Temperature. .

13. **Exam Scores** A student scored 61 on the chemistry final exam, which had a mean of 54 and a standard deviation of 3.5, and she scored 85 on the biology final with a mean of 79 and a standard deviation of 2.5. On which exam did she perform relatively better?

14. **Marathon Run** An amateur male ran a marathon in 3 hours and 51 minutes; the mean time of the male runners was 4 hours and 30 minutes with a standard deviation of 39 minutes. An amateur female ran the same marathon in 4 hours and 40 minutes; the mean time of the female runners was 5 hours and 10 minutes with a standard deviation of 50 minutes. Which marathon runner did relatively better with respect to their gender? *Note:* A lower relative position is better in a marathon run.

Source: Marathon Training Expert.com, What Is the Average Marathon Time?

15. **Graduate Record Exam** A Canadian student applying to an American college wrote the three-part Graduate Record Exam (GRE). The student scores were as follows. In which part of the GRE did the student do relatively better?

Verbal reasoning	Quantitative reasoning	Analytical writing
$X = 593$	$X = 811$	$X = 5.2$
$\overline{X} = 462$	$\overline{X} = 584$	$\overline{X} = 4.0$
$s = 119$	$s = 151$	$s = 0.9$

Source: Educational Testing Services, GRE and Interpreting Your GRE Scores, 2008–2009.

16. **College Room and Board Costs** Room and board costs for selected schools are summarized in this distribution. Find the approximate cost of room and board corresponding to each of the following percentiles.

Costs (in dollars)	Frequency
3000.5–4000.5	5
4000.5–5000.5	6
5000.5–6000.5	18
6000.5–7000.5	24
7000.5–8000.5	19
8000.5–9000.5	8
9000.5–10,000.5	5

 a. 30th percentile
 b. 50th percentile
 c. 75th percentile
 d. 90th percentile

Source: World Almanac.

17. Using the data in Exercise 16, find the approximate percentile rank of each of the following costs.

 a. 5500   *c.* 6500
 b. 7200   *d.* 8300

Section 6.4 Answers

    d. 813 km/h ≈ 79th percentile

    e. 845 km/h ≈ 93rd percentile

23. *a.* $P_{60} = 411$ *c.* $D_4 = P_{40} = 381$

      b. $Q_3 = P_{75} = 415$ *d.* $P_{85} = 427$

25. $P_{70} = 1155$

27. $P_{40} = 2.15$ **29.** $P_{33} = 31$

31. *a.* $Q_1 = 12, Q_2 = 20.5, Q_3 = 32$

      Midquartile $= \frac{12 + 32}{2} = 22$ Interquartile range:
                                            $32 - 12 = 20$

      b. $Q_1 = 62, Q_2 = 94, Q_3 = 99$

      Midquartile $= \frac{62 + 99}{2} = 80.5$ Interquartile range:
                                              $99 - 62 = 37$

Exercises 3–3

1. A *z* score tells how many standard deviations the data value is above or below the mean.

3. A percentile is a relative measure while a percent is an absolute measure of the part to the total.

5. $Q_1 = P_{25}, Q_2 = P_{50}, Q_3 = P_{75}$

7. $D_1 = P_{10}, D_2 = P_{20}, D_3 = P_{30}$, etc.

9. Canada: $z = -0.40$
Italy: $z = 1.47$
United States: $z = -1.91$

11. *a.* 0.75   *b.* -1.25 *c.* 2.25   *d.* -2 *e.* -0.5

13. Chemistry: $z = 2.0$
Biology: $z = 2.4$
The biology exam score is relatively better.

15. Verbal reasoning: $z = 1.1$
Quantitative reasoning: $z = 1.5$
Analytical writing: $z = 1.33$
The quantitative reasoning score is relatively best.

17. *a.* 22nd   *b.* 67th   *c.* 48th   *d.* 88th

19. *a.* 234   *b.* 251   *c.* 263   *d.* 274   *e.* 284

21. a. 611 km/h ≈ 14th percentile
     b. 684 km/h ≈ 41st percentile
     c. 732 km/h ≈ 56th percentile

Ch7 Exercises, Part 1
(Answers at end of Ch7)

10. Give the boundaries of each value.

 a. 42.8 kilometres
 b. 1.6 millilitres
 c. 5.36 grams
 d. 18 kilograms
 e. 13.8 °C
 f. 40 centimetres

11. Name and define the four basic sampling methods.

12. **(ans)** Classify each sample as random, systematic, stratified, or cluster.

 a. In a large school district, all teachers from two buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years.
 b. Every seventh customer entering a shopping mall is asked to select her or his favourite store.
 c. Nursing supervisors are selected using random numbers in order to determine annual salaries.
 d. Every 100th hamburger manufactured is checked to determine its fat content.
 e. Mail carriers of a large city are divided into four groups according to gender (male or female) and according to whether they walk or ride on their routes. Then ten are selected from each group and interviewed to determine whether they have been bitten by a dog in the last year.

13. Give three examples each of nominal, ordinal, interval, and ratio data.

14. For each of these statements, define a population and state how a sample might be obtained.

 a. The average cost of an airline meal is $4.55 (Source: *Everything Has Its Price,* Richard E. Donley, Simon and Schuster).
 b. Some 25% of Canadian children are obese today (Source: *Reader's Digest, Canada*).
 c. Every ten minutes, two people die in car crashes and 170 are injured (Source: National Safety Council estimates).
 d. When older people with mild to moderate hypertension were given mineral salt for six months, the average blood pressure reading dropped by 8 points systolic and 3 points diastolic (Source: *Prevention*).
 e. The average amount spent per gift for Mom on Mother's Day is $25.95 (Source: The Gallup Organization).

15. Select a newspaper or magazine article that involves a statistical study, and write a paper answering these questions.

 a. Is this study descriptive or inferential? Explain your answer.
 b. What are the variables used in the study? In your opinion, what level of measurement was used to obtain the data from the variables?
 c. Does the article define the population? If so, how is it defined? If not, how could it be defined?
 d. Does the article state the sample size and how the sample was obtained? If so, determine the size of the sample and explain how it was selected. If not, suggest a way it could have been obtained.
 e. Explain *in your own words* what procedure (survey, comparison of groups, etc.) might have been used to determine the study's conclusions.
 f. Do you agree or disagree with the conclusions? State your reasons.

16. Information from research studies is sometimes taken out of context. Explain why the claims of these studies might be suspect.

 a. The average salary of the graduates of the class of 1980 is $32,500.
 b. It is estimated that in Podunk there are 27,256 cats.
 c. Only 3% of the men surveyed read *Cosmopolitan* magazine.
 d. Based on a recent mail survey, 85% of the respondents favoured gun control.
 e. A recent study showed that high school dropouts drink more coffee than students who graduated; therefore, coffee dulls the brain.
 f. Since most automobile accidents occur within 24 kms of a person's residence, it is safer to make long trips.

17. Identify each study as being either observational or experimental.

 a. Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After six months, the numbers of respiratory tract infections each group had were compared.
 b. A researcher stood at a busy intersection to see if the colour of the automobile that a person drives is related to running red lights.
 c. A researcher finds that people who are more hostile have higher total cholesterol levels than those who are less hostile.
 d. Subjects are randomly assigned to four groups. Each group is placed on one of four special diets—a low-fat diet, a high-fish diet, a combination of low-fat diet and high-fish diet, and a regular diet. After six months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure.

18. Identify the independent variable(s) and the dependent variable for each of the studies in Exercise 17.

19. For each of the studies in Exercise 17, suggest possible confounding variables.

20. According to a pilot study of 20 people conducted at the University of Minnesota, daily doses of a compound called arabinogalactan over a period of six months resulted in a significant increase in the beneficial lactobacillus species of bacteria. Why can't it be concluded that the compound is beneficial for the majority of people?

Ch 7 Exercises, Part 1

24   Chapter 1 The Nature of Probability and Statistics

21. Comment on the following statement, taken from a magazine advertisement: "In a recent clinical study, Brand ABC [actual brand will not be named] was proved to be 1950% better than creatine!"

22. In an ad for women, the following statement was made: "For every 100 women, 91 have taken the road less travelled." Comment on this statement.

23. In many ads for weight loss products, under the product claims and in small print, the following statement is made: "These results are not typical." What does this say about the product being advertised?

24. In an ad for moisturizing lotion, the following claim is made: " . . . it's the #1 dermatologist-recommended brand." What is misleading about this claim?

25. An ad for an exercise product stated: "Using this product will burn 74% more calories." What is misleading about this statement?

26. "Vitamin E is a proven antioxidant and may help in fighting cancer and heart disease." Is there anything ambiguous about this claim? Explain.

27. "Just 1 capsule of Brand X can provide 24 hours of acid control." (Actual brand will not be named.) What needs to be more clearly defined in this statement?

28. ". . . Male children born to women who smoke during pregnancy run a risk of violent and criminal behaviour that lasts well into adulthood." Can we infer that smoking during pregnancy is responsible for criminal behaviour in people?

29. In the 1980s, a study linked coffee to a higher risk of heart disease and pancreatic cancer. In the early 1990s, studies showed that drinking coffee posed minimal health threats. However, in 1994, a study showed that pregnant women who drank 3 or more cups of tea daily may be at risk for spontaneous abortion. In 1998, a study claimed that women who drank more than a half-cup of caffeinated tea every day may actually increase their fertility. In 1998, a study showed that over a lifetime, a few extra cups of coffee a day can raise blood pressure, heart rate, and stress (Source: "Bottom Line: Is It Good for You? Or Bad?" by Monika Guttman, *USA TODAY Weekend*). Suggest some reasons why these studies appear to be conflicting.

● ● ●

Extending the Concepts

30. Find an article that describes a statistical study, and identify the study as observational or experimental.

31. For the article that you used in Exercise 30, identify the independent variable(s) and dependent variable for the study.

32. For the article that you selected in Exercise 30, suggest some confounding variables that may have an effect on the results of the study.

Statistics Today

Are We Improving Our Diet?—Revisited

Researchers selected a *sample* of 23,699 adults in the United States, using phone numbers selected at *random,* and conducted a *telephone survey.* All respondents were asked six questions:

1. How often do you drink juices such as orange, grapefruit, or tomato?
2. Not counting juice, how often do you eat fruit?
3. How often do you eat green salad?
4. How often do you eat potatoes (not including french fries, fried potatoes, or potato chips)?
5. How often do you eat carrots?
6. Not counting carrots, potatoes, or salad, how many servings of vegetables do you usually eat?

Researchers found that men consumed fewer servings of fruits and vegetables per day (3.3) than women (3.7). Only 20% of the population consumed the recommended five or more daily servings. In addition, they found that youths and less-educated people consumed an even lower amount than the average.

Based on this study, they recommend that greater educational efforts are needed to improve fruit and vegetable consumption by North Americans and to provide environmental and institutional support to encourage increased consumption.

Source: Mary K. Serdula, M.D., et al., "Fruit and Vegetable Intake Among Adults in 16 States: Results of a Brief Telephone Survey," *American Journal of Public Health* 85, no. 2. Copyright by the American Public Health Association; Government of Canada.

Ch7 Exercises, Part 2

More Exercises

Chapter Quiz

● ● ●

Determine whether each statement is true or false. If the statement is false, explain why.

1. Probability is used as a basis for inferential statistics.

2. The height of Sir John A. Macdonald is an example of a variable.

3. The highest level of measurement is the interval level.

4. When the population of college professors is divided into groups according to their rank (instructor, assistant professor, etc.) and then several are selected from each group to make up a sample, the sample is called a cluster sample.

5. The variable *age* is an example of a qualitative variable.

6. The weight of pumpkins is considered to be a continuous variable.

7. The boundary of a value such as 6 centimetres would be 5.9–6.1 centimetres.

Select the best answer.

8. The number of absences per year that a worker has is an example of what type of data?
   a. Nominal
   b. Qualitative
   c. Discrete
   d. Continuous

9. What are the boundaries of 25.6 grams?
   a. 25–26 grams
   b. 25.55–25.65 grams
   c. 25.5–25.7 grams
   d. 20–39 grams

10. A researcher divided subjects into two groups according to gender and then selected members from each group for her sample. What sampling method was the researcher using?
    a. Cluster
    b. Random
    c. Systematic
    d. Stratified

11. Data that can be classified according to colour are measured on what scale?
    a. Nominal
    b. Ratio
    c. Ordinal
    d. Interval

12. A study that involves no researcher intervention is called
    a. An experimental study.
    b. A noninvolvement study.
    c. An observational study.
    d. A quasi-experimental study.

13. A variable that interferes with other variables in the study is called
    a. A confounding variable.
    b. An explanatory variable.
    c. An outcome variable.
    d. An interfering variable.

Use the best answer to complete these statements.

14. Two major branches of statistics are _____ and _____.

15. Two uses of probability are _____ and _____.

16. The group of all subjects under study is called a(n) _____.

17. A group of subjects selected from the group of all subjects under study is called a(n) _____.

18. Three reasons why samples are used in statistics are
    a. _____ b. _____ c. _____.

19. The four basic sampling methods are
    a. _____ b. _____ c. _____ d. _____.

20. A study that uses intact groups when it is not possible to randomly assign participants to the groups is called a(n) _____ study.

21. In a research study, participants should be assigned to groups using _____ methods, if possible.

22. For each statement, decide whether descriptive or inferential statistics is used.
    a. The average life expectancy in New Zealand is 78.49 years. Source: *World Factbook 2004.*
    b. A diet high in fruits and vegetables will lower blood pressure. Source: Institute of Medicine.
    c. The total amount of estimated losses from hurricane Hugo was $4.2 billion. Source: Insurance Service Office.
    d. Researchers stated that the shape of a person's ears is related to the person's aggression. Source: *American Journal of Human Biology.*
    e. In 2013, the number of high school graduates will be 3.2 million students. Source: National Center for Education.

23. Classify each as nominal-level, ordinal-level, interval-level, or ratio-level measurement.
    a. Rating of movies as G, PG, and R.
    b. Number of candy bars sold on a fund drive.
    c. Classification of automobiles as subcompact, compact, standard, and luxury.
    d. Temperatures of hair dryers.
    e. Weights of suitcases on a commercial airline.

Ch 7 Exercises, Part 2

26   **Chapter 1** The Nature of Probability and Statistics

24. Classify each variable as discrete or continuous.
  a. Ages of people working in a large factory.
  b. Number of cups of coffee served at a restaurant.
  c. The amount of a drug injected into a guinea pig.
  d. The time it takes a student to drive to school.

  e. The number of litres of milk sold each day at a grocery store.

25. Give the boundaries of each.
  a. 48 seconds   d. 13.7 kilograms
  b. 0.56 centimetre   e. 7 metres
  c. 9.1 litres

● ● ●

Critical Thinking Challenges

1. A study of the world's busiest airports was conducted by *Airports Council International.* Describe three variables that one could use to determine which airports are the busiest. What *units* would one use to measure these variables? Are these variables categorical, discrete, or continuous?

2. The results of a study published in *Archives of General Psychiatry* stated that male children born to women who smoke during pregnancy run a risk of violent and criminal behaviour that lasts into adulthood. The results of this study were challenged by some people in the media. Give several reasons why the results of this study would be challenged.

3. The results of a study published in *Neurological Research* stated that second-graders who took piano lessons and played a computer math game more readily grasped math problems in fractions and proportions than a similar group who took an English class and played the same math game. What type of inferential study was this? Give several reasons why the piano lessons could improve a student's math ability.

4. A study of 2958 collegiate soccer players showed that in 46 anterior cruciate ligament (ACL) tears, 36 were in women. Calculate the percentages of tears for each gender.

  a. Can it be concluded that female athletes tear their knees more often than male athletes?
  b. Comment on how this study's conclusion might have been reached.

5. Read the article entitled "Anger Can Cause Snap Judgments" and answer the following questions.
  a. Is the study experimental or observational?
  b. What is the independent variable?
  c. What is the dependent variable?
  d. Do you think the sample sizes are large enough to merit the conclusion?
  e. Based on the results of the study, what changes would you recommend to persons to help them reduce their anger?

6. Read the article entitled "Hostile Children Fight Unemployment" and answer the following questions.
  a. Is the study experimental or observational?
  b. What is the independent variable?
  c. What is the dependent variable?
  d. Suggest some confounding variables that may have influenced the results of the study.
  e. Identify the three groups of subjects used in the study.

ANGER CAN CAUSE SNAP JUDGMENTS

Anger can make a normally unbiased person act with prejudice, according to a forthcoming study in the journal *Psychological Science.*

Assistant psychology professors David DeSteno at Northeastern University in Boston and Nilanjana Dasgupta at the University of Massachusetts, Amherst, randomly divided 81 study participants into two groups and assigned them a writing task designed to induce angry, sad or neutral feelings. In a subsequent test to uncover nonconscious associations, angry subjects were quicker to connect negatively charged words —like war, death and vomit—with members of the opposite group—even though the groupings were completely arbitrary.

"These automatic responses guide our behavior when we're not paying attention," says DeSteno, and they can lead to discriminatory acts when there is pressure to make a quick decision. "If you're aware that your emotions might be coloring these gut reactions," he says, "you should take time to consider that possibility and adjust your actions accordingly."

—*Eric Strand*

~~Appendix J~~   Ch 7 Answers

~~Selected Answers~~*

~~Chapter 1~~   Part 1
~~Review Exercises~~

1. Descriptive statistics describes a set of data. Inferential statistics uses a set of data to make predictions about a population.

3. Answers will vary.

5. When the population is large, the researcher saves time and money using samples. Samples are used when the units must be destroyed.

7. *a.* ratio   *e.* ratio   *i.* nominal
   b. ordinal   *f.* ordinal   *j.* ratio
   c. ratio   *g.* ratio
   d. interval   *h.* ratio

8. *a.* quantitative   *d.* quantitative   *g.* qualitative
   b. qualitative   *e.* qualitative
   c. quantitative   *f.* quantitative

9. *a.* discrete   *c.* continuous   *e.* discrete
   b. continuous   *d.* continuous   *f.* continuous

11. Random samples are selected by using chance methods or random numbers. Systematic samples are selected by numbering each subject and selecting every kth number. Stratified samples are selected by dividing the population into groups and selecting from each group. Cluster samples are selected by using intact groups called clusters.

1-in-7

12. *a.* cluster   *c.* random   *e.* stratified
    b. systematic   *d.* systematic

13. Answers will vary.   1-in-100

15. Answers will vary.

17. *a.* experimental   *c.* observational
    b. observational   *d.* experimental

19. Answers will vary. Possible answers include:
    a. overall health of participants, amount of exposure to infected individuals through the workplace or home

b. gender and/or age of driver, time of day
c. diet, general health, heredity factors
d. amount of exercise, heredity factors

21. Claims can be proven only if the entire population is used.

23. Since the results are not typical, the advertisers selected only a few people for whom the product worked extremely well.

25. "74% more calories" than what? No comparison group is stated.

27. What is meant by "24 hours of acid control"?

29. Possible reasons for conflicting results: The amount of caffeine in the coffee or tea or the brewing method.

31. Answers will vary.

~~Chapter Quiz~~   Part 2

1. True   2. False
3. False   4. False
5. False   6. True
7. False   8. *c*
9. *b*   10. *d*
11. *a*   12. *c*
13. *a*   14. descriptive, inferential
15. gambling, insurance   16. population
17. sample
18. *a.* saves time   *c.* use when population is infinite
    b. saves money
19. *a.* random   *c.* cluster
    b. systematic   *d.* stratified
20. quasi-experimental   21. random
22. *a.* descriptive   *d.* inferential
    b. inferential   *e.* inferential
    c. descriptive

*Answers may vary due to rounding or use of technology.

Note: These answers to odd-numbered and selected even-numbered exercises include all quiz answers.

Ch 7 Answers, Part 2

23. *a.* nominal   *d.* interval
    b. ratio   *e.* ratio
    c. ordinal

24. *a.* continuous   *d.* continuous
    b. discrete   *e.* discrete
    c. continuous

25. *a.* 47.5–48.5 seconds
    b. 0.555–0.565 centimetres
    c. 9.05–9.15 quarts
    d. 13.65–13.75 pounds
    e. 6.5–7.5 feet

4.1. COUNTING TECHNIQUES   5

8.1

4.1.1   Exercises

1. How many 2-digit numbers are

   (a) even?

   (b) divisible by 7?

   (c) not divisible by 7?

2. How many 4-digit numbers are

   (a) divisible by 3?

   (b) divisible by 5?

   (c) divisible by 3 and 5?

   (d) divisible by 3 or 5?

   (e) divisible by neither 3 nor 5?

3. A computer system requires a case-sensitive, alpha-numeric password containing 4 or 5 characters. How many possible passwords are there?

4. A computer system requires a case-sensitive, alpha-numeric password containing 5 digits. How many possible passwords are there if

   (a) you can repeat characters?

   (b) you cannot repeat characters?

   (c) you can repeat characters but the first character must be a letter and not a digit?

5. A computer system requires an eight-character, case-sensitive, alpha-numeric passwords.

   (a) How many possible passwords are there?

   (b) How many passwords are there that contain at least one digit?

   (c) How many passwords are there that contain at least one letter?

   (d) How many passwords are there that contain at least one digit and one letter?

6. A computer system requires a case-sensitive, alpha-numeric password containing six characters.

8.1 Exercises

6   *CHAPTER 4. PROBABILITY*

    (a) How many passwords are there that contain no "A"s?

    (b) How many passwords are there that contain no "a"s?

    (c) How many passwords are there that contain no "A"s or "a"s?

7. For homework, Peter has assigned reading pages 25-37 inclusive. How many pages has he asked his class to read?

8. Gilles has assigned for homework all the odd questions from 7 to 89. How many homework questions has he assigned?

9. Canadian postal codes are of the form "letter-number-letter number-letter-number". The first letter shows which province or territory is from, for a total of 13 letters allowable. The remaining letters can be any letter of the alphabet except for O and I. All numbers are allowed. How many possible Canadian postal codes are there?

10. How many days of the week

    (a) contain the letter "t"?

    (b) contain the letter "s"?

    (c) contain the letters "t" and "s"?

    (d) contain the letters "t" or "s"?

11. Pat is writing up systems of equations containing two variables. She will be using lower-case letters for her variables, but doesn't want to use the letters "e", "i", and "o" (for obvious reasons!). How many possible letter combinations does she have to choose from?

12. The mythical Canadian province of Gondor has licence plates of the form "letter-letter number-number-number". Because of an odd superstition, you cannot repeat a letter on the licence plate, but you can repeat a number. How many possible Gondorian licence plates are there?

4.1. COUNTING TECHNIQUES   7

8.1

~~4.1.2~~ Answers

1. First, note that 2-digit numbers run from $10, 11, 12, \ldots 99$.

   (a) The even ones are $10, 12, 14, \ldots 98$. You can do the really short method to count them: __ __ – the first slot can have the digits 1-9 for 9 choices, and the second can only have 2, 4, 6, 8, or 0 for 5 choices. Then the total number is $9 \times 5 = 45$ numbers.

   (b) Unfortunately, you cannot use the above technique for dividing by 7, since 7 doesn't restrict the last digit. Instead, you have to note that the first 2-digit number that's divisible by 7 is 14, the next is 21, then 28, and so on. To find the last digit, you have to count backwards from 99 to find one that's divisible by 7. 99 does not divide evenly by 7, but with your calculator (sigh) you can quickly find that $98 \div 7 = 14$.

   So our sequence is $14, 21, 28, \ldots 98$. This is just $2 \times 7, 3 \times 7, 4 \times 7, \ldots 14 \times 7$. So there are $(\text{last} - \text{first} + 1) = 14 - 2 + 1 = 13$ numbers divisible by 7.

   (c) Total number of 2-digit numbers: last - first + 1 = 90 - 10 + 1 = 90. So the total number of 2-digit numbers **not** divisible by 7 is the total number minus the number that **are** divisible by 7. So, we get $90 - 13 = 77$ for our answer.

2. First, note that 4-digit numbers run from $1000, 1001, 1002, \ldots 9999$.

   (a) The first number that's divisible by 3 is 1002, the next is 1005, then 1008, and so on up to 9999, which also divides evenly by 3.

   So our sequence is $1002, 1005, 1008, \ldots 9999$. This is just $334 \times 3, 335 \times 3, 336 \times 3, \ldots 3333 \times 3$. So there are $(\text{last} - \text{first} + 1) = 3333 - 334 + 1 = 3000$ numbers divisible by 3.

   (b) Numbers that divide evenly by 5 end in either 0 or 5. You can do the really short method to count them: the first slot can have the digits 1-9 for 9 choices, the second and third slots can have 0-9 for 10 choices and the second can only have 0 or 5 for 2 choices. Then the total number is $9 \times 10 \times 10 \times 2 = 1800$ numbers.

   (c) Numbers divisible by 3 **and** 5 must be divisible by 15. Looking at our sequence in a), we can see that 1005 must be the first number, then add 15 to get 1020, etc. Starting from 9999 and

8.1 Answers

    working downwards, we'll see that the first possibility is 9995, which doesn't divide, but 9990 does.

    So our sequence is $1005, 1020, 1035, \ldots 9990$. This is just $67 \times 15, 68 \times 15, 69 \times 15, \ldots 666$. So the total number is $666 - 67 + 1 = 600$.

(d) $n(3 \text{ or } 5) = n(3) + n(5) - n(3 \text{ and } 5) = 3000 + 1800 - 600 = 4200$

(e) total number with 4-digits: $9999 - 1000 + 1 = 9000$ numbers Then the total divisible by neither 3 nor 5 is $9000 - 4200 = 4800$.

3. Case-sensitive, alpha-numeric passwords have $2 \times 26 + 10$ choices for characters, or 62 different possibilities. The number of passwords containing 4 characters is $62 \times 62 \times 62 \times 62 = 14,776,336$. The number of passwords containing 5 characters is $62 \times 62 \times 62 \times 62 \times 62 = 916,132,832$. The total number of passwords is then the sum of these two (since you can't have four **and** five at the same time), $= 930,909,168$.

4. (a) This is the same as in question #3: $916,132,832$.

   (b) If you can't repeat, then you get 62 choices for the first one, 61 for the second, etc., to give $62 \times 61 \times 60 \times 59 \times 58 = 776,520,240$.

   (c) If the first number must be a letter, then you only have 52 possibilities for the first slot: $52 \times 62 \times 62 \times 62 \times 62 = 768,369,472$.

5. (a) You have 62 choices for each slot, so result is $62^8 = 2.18 \times 10^{14}$.

   (b) The number containing no digits is $52^8 = 5.35 \times 10^{13}$. So the number containing at least one digit is $2.18 \times 10^{14} - 5.35 \times 10^{13} = 1.65 \times 10^{14}$.

   (c) The number containing **no** letters is 10^8. So the number containing at least one letter is $2.18 \times 10^{14} - 10^8 = 2.18 \times 10^{14}$ (essentially the same number, since 10^8 is so much smaller).

   (d) The number containing at least one digit and one letter must be the total minus (the number containing no digits plus the number containing no letters). So we get $2.18 \times 10^{14} - 5.35 \times 10^{13} - 10^8 = 1.65 \times 10^{14}$ (very close to the answer to (b) – you'd have to write out a few more decimals to see the difference).

6. (a) If no "A"s are allowed, then we are constrained to 61 choices from our original 62. Then we'll get $61^6 = 5.15 \times 10^{10}$ passwords.

8.1 Answers

4.1. COUNTING TECHNIQUES   9

   (b) This will again give us 61 choices for each character, or $61^6 = 5.15 \times 10^{10}$ passwords.

   (c) Now, we're down to 60 choices, since we can't have "A" or "a". We then get $60^6 = 4.67 \times 10^{10}$ passwords.

7. Peter has assigned 25-37 pages, so $25, 26, 27, \ldots 37$. # pages = last $-$ first $+ 1 = 37 - 25 + 1 = 13$ pages.

8. Gilles has assigned odd questions, so $7, 9, 11, \ldots 89$. It's a bit tricky to do the odd numbers, so I'm going to take all numbers from 7 to 89 and subtract the even numbers.

   total number from 7 to 89: 89 - 7 + 1 = 83

   even numbers: $8, 10, 12, \ldots 88$ is the same as $4 \times 2, 5 \times 2, 6 \times 2, \ldots 44 \times 2$. So we get $44 - 4 + 1 = 41$ even numbers

   odd numbers$= 83 - 41 = 42$ odd numbered questions

9. first letter: 13 choices, second and third letters: 24 choices, all numbers: 10 choices

   So we get $\underline{13} \ \underline{10} \ \underline{24} \ \ \ \underline{10} \ \underline{24} \ \underline{10} = 13 \times 10 \times 24 \times 10 \times 24 \times 10 = 7,488,000$ possible postal codes. (Note that since postal codes reference a geographical area and not a group of people, we're not likely to run out any time soon!)

10. (a) Counting on my fingers, I get that Tuesday, Thursday, and Saturday contain the letter "t" for a total of 3.

   (b) Counting on my fingers, I get that all days except for Monday and Friday have "s" in them for a total of 5.

   (c) I see that all of the days containing "t" also contain "s" for a total of 3.

   (d) Using my counting rules, I get that $n(\text{t or s}) = n(\text{t}) + n(\text{s}) - n(\text{t \& s}) = 3 + 5 - 3 = 5$.

11. I will not be using 3 letters, leaving 23 lower-case letters to choose from. But they have to be different, so I'll get 23×22 choices $= 506$ choices.

12. $\underline{26} \ \underline{25} \ \underline{10} \ \underline{10} \ \underline{10} = 26 \times 25 \times 10^3 = 650,000$.

4.2. PROBABILITY   15

8.2 **4.2.2 Exercises**

1. A fair twelve-sided die is rolled. What is the probability that the roll is

   (a) a 7?

   (b) even?

   (c) greater than 5?

   (d) not a 7?

   (e) a 1 or a 2?

2. Two four-sided dice are rolled. What is the probability that the roll

   (a) results in the same number on both dice?

   (b) results in different numbers on both dice?

   (c) has a sum of 6?

   (d) has at least one die rolling a 3?

3. An individual is presented with three different glasses of soft drink, labeled A, B, and C. He is asked to taste all three and then list them in order of preference. Suppose that the same soft drink has actually been put into all three glasses.

   (a) How many outcomes are there in this experiment? What probability would you assign to each one?

   (b) What is the probability that A is ranked first?

   (c) What is the probability that either B or C is ranked first?

   (d) What is the probability that A is ranked first and B is ranked last?

4. Your ATM/debit card has a four-digit PIN number associated with it. If there are no restrictions on what digits or what order you can pick them, then

   (a) how many PIN numbers are possible?

   (b) what is the probability that someone could guess your PIN randomly?

8.2 Exercises

16   *CHAPTER 4. PROBABILITY*

    (c) if that person saw you input the first two digits when you were at the grocery checkout counter, what are their chances of guessing your PIN correctly now?

Complete the following exercises involving contingency tables.

5. One hundred students each from the Computing Systems Technology program and from the English department were asked who is the greatest fictional wizard ever, with the following results.

	Gandalf	Dumbledore	total
CST	90	10	
English	40	60	
total			

    (a) Calculate $P(G)$.

    (b) Calculate $P(C|G)$.

    (c) Calculate $P(G|C)$.

    (d) Calculate $P(E \text{ or } D)$.

6. A sampling of CST faculty and students were asked what operating system they used on their home computer, with the following results.

	Windows	Linux
faculty	6	2
students	24	8

    (a) What's the probability that a random CST user (faculty or student) will have Linux on their home machine?

    (b) What's the probability that a random CST student will have Linux on their home machine?

    (c) Are the events "student" and "Linux user' independent?

7. One thousand television watchers from BC and Alberta were asked if they watched the Rick Mercer Report on CBC with the following results.

8.2 Exercises

4.2. PROBABILITY   17

	Yes	No
BC	500	500
AB	250	750

(a) What's the probability that one of these people, when selected randomly, is from BC or watches the RMR?

(b) What's the probability that one of these people, when selected randomly, is from BC and watches the RMR?

(c) What's the probability that one of these people, when selected randomly, is from Alberta and does not watch the RMR?

(d) What's the probability that a Rick Mercer watcher is from BC?

(e) What's the probability that a British Columbian watches Rick Mercer?

8. A roving reporter surveyed all of the patrons inside the Starbucks and the Moka House coffee houses in Cook Street Village (it was a slow news day). The beverage each patron was drinking was noted and summarized in the following table.

	coffee	tea	other
Starbucks	45	9	6
Moka House	30	8	2

(a) Are the events "drinking coffee" and "Starbucks" independent?

(b) Are the events "tea" and "Moka House" independent?

9. StatsCan surveyed one hundred Canadians and found that 60 of them exercise regularly, 75 of them eat healthy diets, and 45 of them do both. Complete the following contingency table using the above information

	exercise regularly	don't exercise regularly	total
healthy diet			
unhealthy diet			
total			

(a) If one of these Canadians is selected randomly, what is the probability that this person exercises regularly but does not eat a

8.2 Exercises

18   CHAPTER 4. PROBABILITY

healthy diet?

(b) If one of these Canadians is selected randomly, what is the probability that this person exercises regularly or eats a healthy diet?

(c) Is eating a healthy diet independent of exercising regularly for this sample of Canadians?

4.2. PROBABILITY					19

8.2   4.2.3 **Answers**

1. A fair twelve-sided die is rolled.

   (a) $P(7) = 1/12$ (only one way to get a 7, and there are 12 outcomes)

   (b) Even numbers from 1 to 12: 2, 4, 6, 8, 10, 12, so six possibilities out of 12 outcomes. $P(even) = 6/12 = 1/2$. (Or you could note that exactly half of the outcomes gave an even number to get an even shorter solution.)

   (c) $P(> 5) = P(6$ or 7 or 8 or 9 or 10 or 11 or $12) = 7/12$

   (d) $P(\text{not } 7) = 1 - P(7) = 11/12$

   (e) $P(1$ or $2) = 2/12 = 1/6$

2. I'm going to use the brute force method here and list all possible rolls:

11	12	13	14
21	22	23	24
31	32	33	34
41	42	43	44

   (a) We can see that there are sixteen possibilities in total, and four of them will result in the same number on both dice, so $P(\text{both same}) = 4/16 = 1/4$.

   (b) $P(\text{different}) = 1 - P(\text{same}) = 3/4$.

   (c) $P(\text{sum of } 6) = 3/16$ (need 42, 33, or 24).

   (d) You can count them up to find $P(\text{at least one } 3) = 7/16$. [Or you could say that's $P(\text{at least one } 3) = 1 - P(\text{no 3s})$. And the number of rolls with no 3s is $\underline{3}\ \underline{3} = 9$ possibilities, so then you'd get $1 - 9/16 = 7/16$.]

3. We should note that the order of items matters in this question.

   (a) You can either just calculate or list the possible outcomes: {ABC, ACB, BAC, BCA, CAB, CBA}. Since it's the same soft drink in each glass, the lists should all be equally probable at 1/6 each.

   (b) Only 2 of the 6 outcomes have A first, so $P(A \text{ first}) = 2/6 = 1/3$. Or you could say that if A is first, there are two choices for second

8.2 Answers

20   *CHAPTER 4.   PROBABILITY*

place and one for third (or , if you insist).

(c) If either B or C is ranked first, then A is not. So $P(\text{B or C}) = 1 - \text{P(A)} = 2/3$.

(d) If A is first and B is last, then C is in the middle. $P(ACB) = 1/6$.

4. Your ATM/debit card has a four-digit PIN number associated with it. If there are no restrictions on what digits or what order you can pick them, then

(a) There are 10^4 PIN numbers possible (or 10,000).

(b) If that person is only given one guess (I should have said that!), then their chance is 1/10,000.

(c) If they know the first two numbers, then there are only 10×10 possible PIN numbers left. So their chances are now 1/100.

5. Here is the completed contingency table:

	Gandalf	Dumbledore	total
CST	90	10	100
English	40	60	100
total	130	70	200

(a) $P(G) = 130/200 = 13/20 (or 65\%)$

(b) $P(C|G) = n(CG)/n(G) = 90/130 = 9/13 (which is roughly 69\%)$

(c) $P(G|C) = n(CG)/n(C) = 90/100 = 9/10 (or 90\%)$

(d) $P(E \text{ or D}) = (40 + 60 + 10)/200 = 11/20 (or 55\%)$

8.2 Answers

4.2. PROBABILITY   21

6. Here is the completed contingency table:

	Windows	Linux	total
faculty	6	2	8
students	24	8	40
total	30	10	40

(a) $P(L) = 10/40 = 1/4$ or 25%

(b) $P(L|S) = 8/32 = 1/4$ or 25%

(c) Yes, "student" and "Linux user" are independent because $P(L) = P(L|S) = 1/4$.

7. One thousand television watchers from BC and Alberta were asked if they watched the Rick Mercer Report on CBC with the following results.

	Yes	No	total
BC	500	500	1000
AB	250	750	1000
Total	750	1250	2000

(a) P(BC or Y) = (500 + 500 + 250)/2000 = 5/8 (or 62.5%).

(b) P(BC and Y) = 500/2000 = 1/4 (or 25%).

(c) P(AB and N) = 750/2000 = 3/8 (or 37.5%)

(d) $P(BC|Y)$ = n(BC and Y)/n(Y) = 500/750 = 2/3

(e) $P(Y|BC)$ = n (Y and BC)/n(BC) = 500/1000 = 1/2 (or 50%)

8. Here is the completed contingency table:

	coffee	tea	other	total
Starbucks	45	9	6	60
Moka House	30	8	2	40
Total	75	17	8	100

(a) $P(C) = 75/100 = 3/4 = 75\%$. $P(C|S) = 45/60 = 3/4 = 75\%$. Yes, these events are independent. (You could alternatively cal-

8.2 Answers

culate $P(S) = 60\%$ and $P(S|C) = 60\%$ to reach the same con-
clusion.)

(b) $P(T) = 17/100 = 17\%$. $P(T|M) = 8/40 = 1/5 = 20\%$. As these
are not the same, the events are dependent.

9. (a) Here is the completed contingency table:

	exercise regularly	don't exercise regularly	total
healthy diet	45	30	75
unhealthy diet	15	10	25
total	60	40	100

(b) $P(E \ \overline{H}) = 15/100 = 15\%$.

(c) $P(E \text{ or } H) = (15 + 45 + 30)/100 = 90/100 = 90\%$.

(or $P(E \text{ or } H) = 1 - P(\overline{E} \ \overline{H}) = 1 - 10/100 = 90\%$)

(d) $P(H) = 75/100 = 75\%$. $P(H|E) = 45/60 = 75\%$. Since these
probabilities are the same, eating a healthy diet is independent
of exercising regularly for this sample of Canadians.

Section 8.3: Exercises

4-2 Exercises A: Basic Skills and Concepts

In Exercises 1–4, identify the given random variable as being discrete or continuous.

1. The weight of a randomly selected textbook
2. The cost of a randomly selected textbook
3. The number of eggs a hen lays
4. The amount of milk obtained from a cow

In Exercises 5–12, determine whether a probability distribution is given. In those cases where a probability distribution is not described, identify the requirement that is not satisfied. In those cases where a probability distribution is described, find its mean, variance, and standard deviation.

5. When a household is randomly selected, the probability distribution for the number x of automobiles owned is as described in the accompanying table.

x	$P(x)$
0	0.011
1	0.394
2	0.380
3	0.215

6. If your college hires the next 4 employees without regard to gender, and the pool of applicants is large with an equal number of men and women, then the probability distribution for the number x of women hired is described in the accompanying table.

x	$P(x)$
0	0.0625
1	0.2500
2	0.3750
3	0.2500
4	0.0625

7. Statistics Canada found that English is the mother tongue of 62.6% of the residents in the Greater Toronto Area (GTA). A small survey of 8 residents in one area of the GTA will be taken, so the accompanying table describes the probability distribution for the number of residents (among the 8 randomly selected residents) whose mother tongue is English.

x	$P(x)$
0	0.000
1	0.002
2	0.012
3	0.053
4	0.147
5	0.261

8. In assessing credit risks, a bank investigates the number of credit cards people have. With x representing the number of credit cards adults have, the accompanying table describes the probability distribution for a population of applicants (based on data from Maritz Marketing Research, Inc.).

x	$P(x)$
0	0.26
1	0.16
2	0.12
3	0.09
4	0.07
5	0.09
6	0.07
7	0.14

9. The Willford Printing Company conducted a survey to monitor daily absenteeism at the plant. The accompanying table describes the probability distribution for one department at the plant, where x represents the number of employees absent in the department.

x	$P(x)$
0	0.15
1	0.50
2	0.25
3	0.10

Section 8.3:   Answers

Section 4-2

1. Continuous

3. Discrete

5. Probability distribution with $\mu = 1.799$, $\sigma^2 = 0.613$, $\sigma = 0.782$

7. Not a probability distribution because $\Sigma P(x) \neq 1$.

9. Probability distribution with $\mu = 1.3$, $s^2 = 0.71$, $\sigma = 0.84$

11. Probability distribution with $\mu = 0.5001$, $\sigma^2 = 0.4504$, $\sigma = 0.6711$

13. $-26¢$; $5.26¢$

15. $2.00

17. $\mu = 1.5$, $\sigma^2 = 0.8$, $\sigma = 0.9$; minimum $= -0.3$ and maximum $= 3.3$, but reality dictates that the minimum and maximum are 0 and 3.

19. $\mu = 0.4$, $\sigma^2 = 0.3$, $\sigma = 0.5$

21. a. Yes

    b. No, $\Sigma P(x) > 1$

    c. Yes

    d. Yes

No Exercises for Section 9.1

9.2   Exercises

(Answers at end of ch 9)

9.2

5-2 Exercises A: Basic Skills and Concepts

In Exercises 1–24, assume that the readings on the thermometers are normally distributed with a mean of 0°C and a standard deviation of 1.00°C. A thermometer is randomly selected and tested. In each case, draw a sketch, and find the probability of each reading in degrees.

1. Between 0 and 0.25

2. Between 0 and −0.36

3. Between 0 and 0.89

4. Between 0 and −0.07

9.2 Exercises

5. Between 0 and 3.007

6. Between 0 and 1.96

7. Between 0 and −2.331

8. Between 0 and −1.28

9. Greater than 2.58

10. Less than −1.47

11. Less than −2.09

12. Greater than 0.25

13. Between 1.34 and 2.67

14. Between −1.72 and −0.31

15. Between −2.22 and −1.11

16. Between 0.89 and 1.78

17. Less than 0.08

18. Less than 3.01

19. Greater than −2.29

20. Greater than −1.05

21. Between −1.99 and 2.01

22. Between −0.07 and 2.19

23. Between −1.00 and 4.00

24. Between −5.00 and 2.00

In Exercises 25–28, assume that the readings on the thermometers are normally distributed with a mean of 0°C and a standard deviation of 1.00°C. Find the indicated probability, where z is the reading in degrees.

25. $P(z > 2.33)$

26. $P(2.00 < z < 2.50)$

27. $P(−3.00 < z < 2.00)$

28. $P(z < −1.44)$

In Exercises 29–36, assume that the readings on the thermometers are normally distributed with a mean of 0°C and a standard deviation of 1.00°C. A thermometer is randomly selected and tested. In each case, draw a sketch, and find the temperature reading corresponding to the given information.

29. Find P_{90}, the 90th percentile. This is the temperature reading separating the bottom 90% from the top 10%.

30. Find P_{30}, the 30th percentile.

31. Find Q_1, the temperature reading that is the first quartile.

32. Find D_1, the temperature reading that is the first decile.

33. If 4% of the thermometers are rejected because they have readings that are too high, but all other thermometers are acceptable, find the reading that separates the rejected thermometers from the others.

34. If 8% of the thermometers are rejected because they have readings that are too low, but all other thermometers are acceptable, find the reading that separates the rejected thermometers from the others.

35. A quality control analyst wants to examine thermometers that give readings in the bottom 2%. What reading separates the bottom 2% from the others?

36. If 2.5% of the thermometers are rejected because they have readings that are too high and another 2.5% are rejected because they have readings that are too low, find the two readings that are cutoff values separating the rejected thermometers from the others.

9.2 Exercises

9.2

〽️ Exercises B: Beyond the Basics

37. Assume that z scores are normally distributed with a mean of 0 and a standard deviation of 1.
    a. If $P(0 < z < a) = 0.3212$, find a.
    b. If $P(-b < z < b) = 0.3182$, find b.
    c. If $P(z > c) = 0.2358$, find c.
    d. If $P(z > d) = 0.7517$, find d.
    e. If $P(z < e) = 0.4090$, find e.

38. For a standard normal distribution, find the percentage of data that are
    a. within 1 standard deviation of the mean
    b. within 1.96 standard deviations of the mean
    c. between $\mu - 3\sigma$ and $\mu + 3\sigma$
    d. between 1 standard deviation below the mean and 2 standard deviations above the mean
    e. Suppose it turns out that the distribution is not exactly normal, but is positively skewed. How does this affect your answers to parts (a), (b), and (d) of this exercise?

39. In a manufacturing plant that makes boxes, the width of a certain type of box is normally distributed. The probability that the width is less than 23.9708 cm is 0.0721 and the probability that the width is more than 24.0404 cm is 0.0217. Find the mean and standard deviation of the box width.

40. In a certain region, annual household incomes are normally distributed. The middle 95% of the incomes are between $72,684 and $78,564. Find the mean and standard deviation of the annual household incomes for this region.

5-3 Normal Distributions: Finding Probabilities

Although Section 5–2 introduced important methods for dealing with normal distributions, the examples and exercises included in that section are generally unrealistic because most normally distributed populations have a nonzero mean, a standard deviation different from 1, or both. In this section we include many real and important nonstandard normal distributions. The basic principle we will be explaining in this section is the following:

> If we convert values to standard scores using Formula 5-2, then procedures for working with all normal distributions are the same as for the standard normal distribution.

Formula 5-2
$$z = \frac{x - \mu}{\sigma}$$

9.3 Exercises
(Answers at end of Ch 9)

9.3

5-3 **Exercises A: Basic Skills and Concepts**

In Exercises 1–6, assume that the heights of female students are normally distributed with a mean given by μ = 64.2 in. and a standard deviation given by σ = 2.6 in. (based on data from a survey of college students). Also assume that a female student is randomly selected. Draw a graph, and find the indicated probability.

1. $P(64.2 \text{ in.} < x < 65.0 \text{ in.})$ 2. $P(x < 70.0 \text{ in.})$

3. $P(x > 58.1 \text{ in.})$ 4. $P(59.1 \text{ in.} < x < 66.6 \text{ in.})$

5. A fashion agency is looking for females between 65.5 in. and 68.0 in. tall to work as models. Find the probability that a randomly selected female student meets the height requirements to be a model.

6. The Beanstalk Club, a social organization for tall people, has a requirement that women must be at least 70 in. (or 5 ft 10 in.) tall. Suppose you are trying to decide whether to open a branch of the Beanstalk Club at your college with 500 female students.
   a. Find the percentage of female students who are eligible for membership because they meet the minimum height requirement of 70 in.
   b. Among the 500 female students in your college, how many would be eligible for Beanstalk Club membership?
   c. Will you open a branch of the Beanstalk Club?

7. Replacement times for TV sets are normally distributed with a mean of 8.2 years and a standard deviation of 1.1 years (based on data from "Getting Things Fixed," *Consumer Reports*). Find the probability that a randomly selected TV set will have a replacement time of less than 7.0 years.

8. Replacement times for CD players are normally distributed with a mean of 7.1 years and a standard deviation of 1.4 years (based on data from "Getting Things Fixed," *Consumer Reports*). Find the probability that a randomly selected CD player will have a replacement time of less than 8.0 years.

9. Assume that the heights of soldiers in the Canadian Armed Forces are normally distributed with a mean height of 70.3 in. and a standard deviation of 3.4 in. Find the probability of one soldier who is selected at random having a height of 77.0 in. or greater.

10. Based on the sample results in Data Set 18 of Appendix B, assume that human body temperatures are normally distributed with a mean of 36.4°C and a standard deviation of 0.62°C. If we define a fever to be a body temperature above 37.8°C, what percentage of normal and healthy persons would be considered to have a fever? Does this percentage suggest that a cutoff of 37.8°C is appropriate?

9.3 Exercises

11. One classic use of the normal distribution is inspired by a letter to *Dear Abby* in which a wife claimed to have given birth 308 days after a brief visit from her husband, who was serving in the Navy. The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days. Given this information, find the probability of a pregnancy lasting 308 days or longer. What does the result suggest?

12. Lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days. If we stipulate that a baby is *premature* if born at least three weeks early, what percentage of babies are born prematurely? Why would this information be useful to hospital administrators?

13. Based on daily summaries from a Calgary observatory (for January to November 2000), the mean daily counting rates for cosmic rays are approximately normally distributed, with a mean equal to 3465.5 and a standard deviation of 127.7. If one day is randomly selected, what is the probability that the day's observed mean counting rate is at least 3248?

14. According to the International Mass Retail Association, girls aged 13 to 17 spend an average of $31.20 on shopping trips in a month. Assume that the amounts are normally distributed with a standard deviation of $8.27.

    If a girl in that age category is randomly selected, what is the probability that she spends between $35.00 and $40.00 in one month? Does the assumption of a normal distribution seem plausible for this population?

15. IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. Mensa is an organization for people with high IQs, and eligibility requires an IQ above 131.5.
    a. If someone is randomly selected, find the probability that he or she meets the Mensa requirement.
    b. In a typical region of 75,000 people, how many are eligible for Mensa?

16. An IBM subcontractor was hired to make ceramic substrates that are used to distribute power and signals to and from computer silicon chips. Specifications require resistance between 1.500 ohms and 2.500 ohms, but the population has normally distributed resistances with a mean of 1.978 ohms and a standard deviation of 0.172 ohms. What percentage of the ceramic substrates will not meet the manufacturer's specifications? Does this manufacturing process appear to be working well?

17. The average household expenditure in Canada on postsecondary books is $53.00, with a standard deviation of $18.61 (based on a study by Statistics Canada on family expenditures). If a household is selected at random, find the probability that its expenditure on postsecondary books is between $60.00 and $70.00. Do you expect that the household expenditure on books is normally distributed?

9.3 Exercises

18. Measurements of human skulls from different epochs are analyzed to determine whether they change over time. The maximum breadth is measured for skulls from Egyptian males who lived around 3300 BCE. Results show that those breadths are normally distributed with a mean of 132.6 mm and a standard deviation of 5.4 mm (based on data from *Ancient Races of the Thebaid* by Thomson and Randall-Maciver). An archeologist discovers a male Egyptian skull and a field measurement reveals a maximum breadth of 119 mm. Find the probability of getting a value of 119 or less if a skull is randomly selected from the period around 3300 BCE. Is the newly found skull likely to come from that era?

19. According to a national health survey, the serum cholesterol levels of men aged 18 to 24 are normally distributed with a mean and standard deviation (in mg/100mL) of 178.1 and 40.7, respectively. One criterion for identifying risk of coronary disease is a cholesterol level above 300. If a man aged 18–24 is randomly selected, find the probability that his serum cholesterol level is above 300. Does this probability warrant serious concern?

20. Some vending machines are designed so that their owners can adjust the weights of the quarters that are accepted. If many counterfeit coins are found, adjustments are made to reject more coins, with the effect that most of the counterfeit coins are rejected along with many legal coins. Assume that quarters have weights that are normally distributed with a mean of 5.67 g and a standard deviation of 0.070 g. If a vending machine is adjusted to reject quarters weighing less than 5.50 g or more than 5.80 g, what is the percentage of legal quarters that are rejected?

5-3 Exercises B: Beyond the Basics

In Exercises 21–23, refer to the indicated data set in Appendix B.
  a. *Construct a histogram to determine whether the data set has a normal distribution.*
  b. *Find the sample mean and sample standard deviation s.*
  c. *Use the sample mean as an estimate of the population mean μ, use the sample standard deviation as an estimate of the population standard deviation σ, and use the methods of this section to find the indicated probability.*

21. Use the combined list of 100 weights of M&M plain candies listed in Data Set 11, and estimate the probability of randomly selecting one M&M candy and getting one with a weight greater than 1.000 g.

22. Use the total weights of discarded garbage in Data Set 1, and estimate the probability of randomly selecting a household that discards more than 20.0 lb of garbage in a week.

9·4 Exercises
(Answers at end of Ch 9)

9.4

~~5-4~~ **Exercises A: Basic Skills and Concepts**

In Exercises 1–4, assume that female college students have heights that are normally distributed with a mean of 64.2 in. and a standard deviation of 2.6 in. Find the height for the given percentile.

1. P_{85}

2. P_{66}

3. P_{15}

4. P_{35}

5. Replacement times for TV sets are normally distributed with a mean of 8.2 years and a standard deviation of 1.1 years (based on data from "Getting Things Fixed," *Consumer Reports*).
   a. Find the replacement time that separates the top 20% from the bottom 80%.
   b. Find the probability that a randomly selected TV will have a replacement time of less than 5.0 years.
   c. If you want to provide a warranty so that only 1% of the TV sets will be replaced before the warranty expires, what length of time would you recommend for the warranty?

6. Replacement times for CD players are normally distributed with a mean of 7.1 years and a standard deviation of 1.4 years (based on data from "Getting Things Fixed," *Consumer Reports*). Find P_{55}, which is the replacement time separating the top 45% from the bottom 55%.

7. Weights of paper discarded by households each week are normally distributed with a mean of 4.3 kg and a standard deviation of 1.9 kg. Find P_{33}, which is the weight that separates the bottom 33% from the top 67%.

8. Based on the sample results in Data Set 18 of Appendix B, assume that human body temperatures are normally distributed with a mean of 36.4°C and a standard deviation of 0.62°C. What two temperature levels separate the bottom 2% and the top 2%? Could these values serve as reasonable limits that could be used to identify people who are likely to be ill?

9. The durations of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days. If we stipulate that a baby is *premature* if the length of pregnancy is in the lowest 4%, find the duration that separates premature babies from those who are not premature.

10. According to the Opinion Research Corporation, men spend an average of 11.4 min in the shower. Assume that the times are normally distributed with a standard deviation of 1.8 min. Find the values of the quartiles Q_1 and Q_3.

11. IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. If we define a genius to be someone in the top 1% of IQ scores, find the score separating geniuses from the rest of us. Are there any jobs where this score could reasonably be used as one criterion for employment?

9.4 Exercises

12. A subcontractor manufactures ceramic substrates for IBM. These devices have resistances that are normally distributed with a mean of 1.978 ohms and a standard deviation of 0.172 ohms. If the required specifications are to be modified so that 3% of the devices are rejected because their resistances are too low and another 3% are rejected because their resistances are too high, find the cutoff values for the acceptable devices.

13. Scores obtained from the Law School Admission Test (LSAT) are normally distributed with a mean score of 550 and a standard deviation of 110.
    a. If a test score is selected at random, find the probability that it is less than 750.
    b. If the top 16% of the test scores are usually good enough for admission to law school, find the cutoff score for gaining admission.

14. Measurements of human skulls from different epochs are analyzed to determine whether they change over time. The maximum breadth is measured for skulls from Egyptian males who lived around 3300 BCE. Results show that those breadths are normally distributed with a mean of 132.6 mm and a standard deviation of 5.4 mm (based on data from *Ancient Races of the Thebaid* by Thomson and Randall-Maciver).
    a. Find the probability of getting a value greater than 140 mm if a skull is randomly selected from the period of around 3300 BCE.
    b. Find the value that is D_2, the second decile.

15. Based on daily summaries from a Calgary observatory over a period of ten months, the mean daily counting rates for cosmic rays are approximately normally distributed, with a mean equal to 3465.5 and a standard deviation of 127.7. Find the sixth decile D_6. What is the mean daily counting rate that separates the lowest 60% from the highest 40%?

16. Quarters have weights that are normally distributed with a mean of 5.67 g and a standard deviation of 0.070 g.
    a. If a vending machine is adjusted to reject quarters weighing less than 5.53 g or more than 5.81 g, what is the percentage of legal quarters that are rejected?
    b. Find the weights of accepted legal quarters if the machine is readjusted so that the lightest 1.5% are rejected and the heaviest 1.5% are rejected.
    c. If your quarter is rejected from a machine that is set to reject the upper 1.5% and the lower 1.5% of coins, by weight, is it a waste of time to re-insert your coin?

5-4 Exercises B: Beyond the Basics

17. The construction of a histogram for a data set reveals that the distribution is approximately normal and the boxplot is constructed with these quartiles: $Q_1 = 62$, $Q_2 = 70$, $Q_3 = 78$. Estimate the standard deviation.

No Exercises for 9.5

Ch9 Answers on next page →

Ch9 Answers

9. $P(\mu - 2\sigma) < X < \mu + 2\sigma) = 4\sigma/(b - a)$
   Since $\sigma = (b - a)/\sqrt{12}$, we get $P(\mu - 2\sigma < X < \mu + 2\sigma) = 4/\sqrt{12} > 1$.
   Since the probability cannot exceed 100%, we can deduce that 100% of a uniform distribution lies within 2 standard deviations of the mean, regardless of the values of a and b.

Section 9.2

1. 0.0987
3. 0.3133
5. 0.4987
7. 0.4901
9. 0.0049
11. 0.0183
13. 0.0863
15. 0.1203
17. 0.5319
19. 0.9890
21. 0.9545
23. 0.8412
25. 0.0099
27. 0.9759
29. 1.28°
31. −0.67°
33. 1.75°
35. −2.05°
37. a. 0.92
   b. 0.41
   c. 0.72
   d. −0.68
   e. −0.23
39. $\mu = 24$; $\sigma = 0.02$

Section 9.3

1. 0.1217
3. 0.9906
5. 0.2364
7. 0.1379
9. 0.0244
11. 0.0038; either a very rare event has occurred or the husband is not the father.

13. 0.9554
15. a. 0.0179
   b. 1343
17. 0.1706. This answer assumes a normal distribution, but the expenditure distribution is more likely to be bimodal: Many households will spend nothing on postsecondary textbooks (because no one is attending postsecondary institutions); and the minority who are buying textbooks will spend a great deal.
19. 0.0013. No, it is rare for a person in this age group to have a seriously elevated serum cholesterol level.
21. a. Normal distribution
   b. $\bar{x} = 0.9147, s = 0.0369$
   c. 0.0104
23. a. Normal distribution
   b. $\bar{x} = 35.66, s = 9.35$
   c. 0.271

Section 9.4

1. 66.9 in.
3. 61.5 in.
5. a. 9.1 years
   b. 0.0018
   c. 5.6 years.
7. 3.5 kg
9. 242 days
11. 135. Perhaps the employees of a "think tank" company could be held to this standard.
13. a. 0.9656
   b. 658.9
15. 3497.4
17. 11.9
19. 131.6; 782

No Exercises for Section 10.1

Multiplying by -1, one gets

$$\overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Reversing the inequality, one gets the formula for the confidence interval:

$$\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Applying the Concepts 7–2

Making Decisions with Confidence Intervals

Assume you work for Kimberly Clark Corporation, the makers of Kleenex. The job you are presently working on requires you to decide how many Kleenexes are to be put in the new automobile glove compartment boxes. Complete the following.

1. How will you decide on a reasonable number of Kleenexes to put in the boxes?
2. When do people usually need Kleenexes?
3. What type of data collection technique would you use?
4. Assume you found out that from your sample of 85 people, on average about 57 Kleenexes are used throughout the duration of a cold, with a standard deviation of 15. Use a confidence interval to help you decide how many Kleenexes will go in the boxes.
5. Explain how you decided on how many Kleenexes will go in the boxes.

See page 331 for the answers.

Exercises   for   Section   10.2

Exercises 7–2

1. What is the difference between a point estimate and an interval estimate of a parameter? Which is better? Why?

2. What information is necessary to calculate a confidence interval?

3. What is the margin of error?

4. What is meant by the 95% confidence interval of the mean?

5. What are three properties of a good estimator?

6. What statistic best estimates μ?

7. What is necessary to determine the sample size?

8. When one is determining the sample size for a confidence interval, is the size of the population relevant?

9. Find each.

   a. $z_{\alpha/2}$ for the 99% confidence interval
   b. $z_{\alpha/2}$ for the 98% confidence interval
   c. $z_{\alpha/2}$ for the 95% confidence interval
   d. $z_{\alpha/2}$ for the 90% confidence interval
   e. $z_{\alpha/2}$ for the 94% confidence interval

10. Find the 95% confidence interval for the mean paid attendance at the Major League Baseball All-Star games. A random sample of the paid attendances is shown.

47,596	68,751	5,838
69,831	28,843	53,107
31,391	48,829	50,706
62,892	55,105	63,974
56,674	38,362	51,549
31,938	31,851	56,088
34,906	38,359	72,086
34,009	50,850	43,801
46,127	49,926	54,960
32,785	48,321	49,671

Source: *Time Almanac.*

10.2 Exercises

11. A sample of the reading scores of 35 Grade 5 students has a mean of 82. The standard deviation of the sample is 15.

   a. Find the best point estimate of the mean.

   b. Find the 95% confidence interval of the mean reading scores of all Grade 5 students.

   c. Find the 99% confidence interval of the mean reading scores of all Grade 5 students.

   d. Which interval is larger? Explain why.

12. Find the 90% confidence interval of the population mean for the number of detached house sales in Toronto districts over a one-year period. A random sample of 40 districts is shown.

941	573	2864	739
920	759	889	928
1461	799	667	1991
988	718	1280	1137
1278	921	272	624
546	1106	1019	913
1285	535	1463	1377
910	1208	435	2124
1145	538	855	888
650	1105	455	2306

Source: Toronto Real Estate Board.

13. A study of 40 English composition professors showed that they spent, on average, 12.6 minutes correcting a student's term paper.

   a. Find the best point estimate of the mean.

   b. Find the 90% confidence interval of the mean time for all composition papers when $\sigma = 2.5$ minutes.

   c. If a professor stated that he spent, on average, 30 minutes correcting a term paper, what would be your reaction?

14. A study of 35 golfers showed that their average score on a particular course was 92. The standard deviation of the sample is 5.

   a. Find the best point estimate of the mean.

   b. Find the 95% confidence interval of the mean score for all golfers.

   c. Find the 95% confidence interval of the mean score if a sample of 60 golfers is used instead of a sample of 35.

   d. Which interval is smaller? Explain why.

15. A survey of individuals who passed the seven exams and obtained the rank of Fellow in the actuarial field finds the average salary to be $150,000. If the standard deviation for the sample of 35 Fellows was

$15,000, construct a 95% confidence interval for all Fellows.

Source: www.BeAnActuary.org.

16. A survey of 30 gas stations randomly selected nationwide on Canada Day 2006 indicate the following prices (cents per litre) of regular gasoline fuel. Estimate the average price per litre of fuel with 90% confidence.

117.8	96.4	99.9	96.9	111.9	110.2
109.9	100.4	109.5	96.6	104.5	114.9
114.9	99.9	108.8	103.2	95.8	110.7
112.9	109.5	109.9	103.2	105.4	122.8
108.2	108.8	104.2	96.9	107.4	114.9

Source: MJ Ervin & Associates.

17. A study of 415 kindergarten students showed that they have seen on average 5000 hours of television. If the sample standard deviation is 900, find the 95% confidence level of the mean for all students. If a parent claimed that his children watched 4000 hours, would the claim be believable?

Source: U.S. Department of Education.

18. A random sample of 76 four-year-olds attending day-care centres showed that the yearly tuition averaged $3648. The standard deviation of the sample was $630, and the sample size was 50. Find the 90% confidence interval of the true mean. If a day-care centre were starting up and wanted to keep tuition low, what would be a reasonable amount to charge?

19. Noise levels at various area urban hospitals were measured in decibels. The mean of the noise levels in 84 corridors was 61.2 decibels, and the standard deviation was 7.9. Find the 95% confidence interval of the true mean.

Source: M. Bayo, A. Garcia, and A. Garcia, "Noise Levels in an Urban Hospital and Workers' Subjective Responses," *Archives of Environmental Health* 50, no. 3, p. 249 (May–June 1995). Reprinted with permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 Eighteenth St. N.W., Washington, D.C. 20036-1802. Copyright © 1995.

20. The growing seasons for a random sample of 35 United States cities were recorded, yielding a sample mean of 190.7 days and a sample standard deviation of 54.2 days. Estimate the true mean population of the growing season with 95% confidence.

Source: *The Old Farmer's Almanac.*

21. A university dean of students wishes to estimate the average number of hours students spend doing homework per week. The standard deviation from a previous study is 6.2 hours. How large a sample must be selected if he wants to be 99% confident of finding whether the true mean differs from the sample mean by 1.5 hours?

10.2 Exercises

22. In the hospital study cited in Exercise 19, the mean noise level in the 171 ward areas was 58.0 decibels, and the standard deviation was 4.8. Find the 90% confidence interval of the true mean.

    Source: M. Bayo, A. Garcia, and A. Garcia, "Noise Levels in an Urban Hospital and Workers' Subjective Responses," *Archives of Environmental Health* 50, no. 3, p. 249 (May–June 1995). Reprinted with permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 Eighteenth St. N.W., Washington, D.C. 20036-1802. Copyright © 1995.

23. An insurance company is trying to estimate the average number of sick days that full-time food service workers use per year. A pilot study found the standard deviation to be 2.5 days. How large a sample must be selected if the company wants to be 95% confident of getting an interval that contains the true mean with a maximum error of 1 day?

24. A pizza shop owner wishes to find the 95% confidence interval of the true mean cost of a large plain pizza. How large should the sample be if she wishes to be accurate to within $0.15? A previous study showed that the standard deviation of the price was $0.26.

25. A researcher is interested in estimating the average monthly salary of sports reporters in a large city. He wants to be 90% confident that his estimate is correct. If the standard deviation is $1100, how large a sample is needed to get the desired information and to be accurate to within $150?

15. $145,030 < \mu < $154,970

17. $4913 < \mu < 5087$; 4000 hours does not seem reasonable since it is outside this interval.

19. $59.5 < \mu < 62.9$　　　21. 114

23. 25　　　　　　　　　　　25. 147

10.2 Answers

Chapter 7
Exercises 7–2

1. A point estimate of a parameter specifies a specific value such as $\mu = 87$, whereas an interval estimate specifies a range of values for the parameter such as $84 < \mu < 90$. The advantage of an interval estimate is that a specific confidence level (say 95%) can be selected, and one can be 95% confident that the parameter being estimated lies in the interval.

3. The maximum error of estimate is the likely range of values to the right or left of the statistic in which may contain the parameter.

5. A good estimator should be unbiased, consistent, and relatively efficient.

7. To determine sample size, the maximum error of estimate and the degree of confidence must be specified and the population standard deviation must be known.

9. *a.* 2.575　　　*c.* 1.96　　　*e.* 1.88
   b. 2.33　　　*d.* 1.645

11. *a.* 82　　*b.* $77 < \mu < 87$　　　*c.* $75 < \mu < 89$
    d. The 99% confidence interval is larger because the confidence level is larger.

13. *a.* 12.6　　*b.* $11.9 < \mu < 13.3$
    c. It would be highly unlikely since this is far larger than 13.3 minutes.