1.1 Give a specific example of data that might be gathered from each of the following business disciplines: accounting, finance, human resources, marketing, information systems, production, and management. An example in the marketing area might be "number of sales per month by each salesperson."

1.2 State examples of data that can be gathered for decision-making purposes from each of the following industries: manufacturing, insurance, travel, retailing, communications, computing, agriculture, banking, and healthcare. An example in the travel industry might be the cost of business travel per day in various European cities.

1.3 Give an example of descriptive statistics in the recorded music industry. Give an example of how inferential statistics could be used in the recorded music industry. Compare the two examples. What makes them different?

1.4 Suppose you are an operations manager for a plant that manufactures batteries. Give an example of how you could use descriptive statistics to make better managerial decisions. Give an example of how you could use inferential statistics to make better managerial decisions.

1.5 Classify each of the following as nominal, ordinal, interval, or ratio level data.
   a. The time required to produce each tire on an assembly line
   b. The number of quarts of milk a family drinks in a month
   c. The ranking of four machines in your plant after they have been designated as excellent, good, satisfactory, and poor
   d. The telephone area code of clients in the United States
   e. The age of each of your employees
   f. The dollar sales at the local pizza house each month
   g. An employee's ID number
   h. The response time of an emergency unit

1.6 Classify each of the following as nominal, ordinal, interval, or ratio level data.
   a. The ranking of a company by Fortune 500
   b. The number of tickets sold at a movie theater on any given night
   c. The identification number on a questionnaire
   d. Per capita income
   e. The trade balance in dollars
f. Socioeconomic class (low, middle, upper)
g. Profit loss in dollars
h. A company’s tax ID
i. The Standard & Poor’s bond ratings of cities based on the following scales.

<table>
<thead>
<tr>
<th>RATING</th>
<th>GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest quality</td>
<td>AAA</td>
</tr>
<tr>
<td>High quality</td>
<td>AA</td>
</tr>
<tr>
<td>Upper medium quality</td>
<td>A</td>
</tr>
<tr>
<td>Medium quality</td>
<td>BBB</td>
</tr>
<tr>
<td>Somewhat speculative</td>
<td>BB</td>
</tr>
<tr>
<td>Low quality, speculative</td>
<td>B</td>
</tr>
<tr>
<td>Low grade, default possible</td>
<td>CCC</td>
</tr>
<tr>
<td>Low grade, partial recovery</td>
<td>CC</td>
</tr>
<tr>
<td>Default, recovery unlikely</td>
<td>C</td>
</tr>
</tbody>
</table>

1.7 The Rathburn Manufacturing Company makes electric wiring, which it sells to contractors in the construction industry. Approximately 900 electric contractors purchase wire from Rathburn annually. Rathburn’s director of marketing wants to determine electric contractors’ satisfaction with Rathburn’s wire. He developed a questionnaire that yields a satisfaction score of between 10 and 50 for participant responses. A random sample of 35 of the 900 contractors is asked to complete a satisfaction survey. The satisfaction scores for the 35 participants are averaged to produce a mean satisfaction score.

a. What is the population for this study?
b. What is the sample for this study?
c. What is the statistic for this study?
d. What would be a parameter for this study?
Supplementary Problems

Calculating the Statistics

2.14 For the following data, construct a frequency distribution with six classes.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>23</td>
<td>35</td>
<td>18</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>51</td>
<td>47</td>
<td>29</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>43</td>
<td>29</td>
<td>36</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>41</td>
<td>19</td>
<td>23</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>42</td>
<td>52</td>
<td>29</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>46</td>
<td>33</td>
<td>28</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

2.15 For each class interval of the frequency distribution given, determine the class midpoint, the relative frequency, and the cumulative frequency.

<table>
<thead>
<tr>
<th>CLASS INTERVAL</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–under 25</td>
<td>17</td>
</tr>
<tr>
<td>25–under 30</td>
<td>20</td>
</tr>
<tr>
<td>30–under 35</td>
<td>16</td>
</tr>
<tr>
<td>35–under 40</td>
<td>15</td>
</tr>
<tr>
<td>40–under 45</td>
<td>8</td>
</tr>
<tr>
<td>45–under 50</td>
<td>6</td>
</tr>
</tbody>
</table>
2.16 Construct a histogram, a frequency polygon, and an ogive for the following frequency distribution.

<table>
<thead>
<tr>
<th>CLASS INTERVAL</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>50–under 60</td>
<td>13</td>
</tr>
<tr>
<td>60–under 70</td>
<td>27</td>
</tr>
<tr>
<td>70–under 80</td>
<td>43</td>
</tr>
<tr>
<td>80–under 90</td>
<td>31</td>
</tr>
<tr>
<td>90–under 100</td>
<td>9</td>
</tr>
</tbody>
</table>

2.17 Construct a pie chart from the following data.

<table>
<thead>
<tr>
<th>LABEL</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>55</td>
</tr>
<tr>
<td>B</td>
<td>121</td>
</tr>
<tr>
<td>C</td>
<td>83</td>
</tr>
<tr>
<td>D</td>
<td>46</td>
</tr>
</tbody>
</table>

2.18 Construct a stem and leaf plot for the following data. Let the leaf contain one digit.

312 324 289 335 298
314 309 294 326 317
290 311 317 301 316
306 286 308 284 324

2.19 The Whitcomb Company manufactures a metal ring for industrial engines that usually weighs about 50 oz. A random sample of 50 of these metal rings produced the following weights (in ounces).

<table>
<thead>
<tr>
<th>51</th>
<th>53</th>
<th>56</th>
<th>50</th>
<th>44</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>53</td>
<td>42</td>
<td>57</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>41</td>
<td>44</td>
<td>52</td>
<td>56</td>
<td>50</td>
<td>57</td>
</tr>
<tr>
<td>44</td>
<td>46</td>
<td>41</td>
<td>52</td>
<td>69</td>
<td>53</td>
</tr>
<tr>
<td>57</td>
<td>51</td>
<td>54</td>
<td>63</td>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td>47</td>
<td>52</td>
<td>53</td>
<td>46</td>
<td>36</td>
<td>58</td>
</tr>
<tr>
<td>51</td>
<td>38</td>
<td>49</td>
<td>50</td>
<td>62</td>
<td>39</td>
</tr>
<tr>
<td>44</td>
<td>55</td>
<td>43</td>
<td>52</td>
<td>43</td>
<td>42</td>
</tr>
<tr>
<td>57</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Construct a frequency distribution for these data using eight classes. What can you observe from the frequency distribution about the data?

2.20 A northwestern distribution company surveyed 53 of its midlevel managers. The survey obtained the ages of these managers, which later were organized into the frequency distribution shown. Determine the class midpoint, relative frequency, and cumulative frequency for these data.

<table>
<thead>
<tr>
<th>CLASS INTERVAL</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–under 25</td>
<td>8</td>
</tr>
<tr>
<td>25–under 30</td>
<td>6</td>
</tr>
<tr>
<td>30–under 35</td>
<td>5</td>
</tr>
<tr>
<td>35–under 40</td>
<td>12</td>
</tr>
<tr>
<td>40–under 45</td>
<td>15</td>
</tr>
<tr>
<td>45–under 50</td>
<td>7</td>
</tr>
</tbody>
</table>

2.21 The following data are shaped roughly like a normal distribution (discussed in Chapter 6).

<table>
<thead>
<tr>
<th>61.4</th>
<th>27.3</th>
<th>26.4</th>
<th>37.4</th>
<th>30.4</th>
<th>47.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>63.9</td>
<td>46.8</td>
<td>67.9</td>
<td>19.1</td>
<td>81.6</td>
<td>47.9</td>
</tr>
<tr>
<td>73.4</td>
<td>54.6</td>
<td>65.1</td>
<td>53.3</td>
<td>71.6</td>
<td>58.6</td>
</tr>
<tr>
<td>57.3</td>
<td>87.8</td>
<td>71.1</td>
<td>74.1</td>
<td>48.9</td>
<td>60.2</td>
</tr>
<tr>
<td>54.8</td>
<td>60.5</td>
<td>32.5</td>
<td>61.7</td>
<td>55.1</td>
<td>48.2</td>
</tr>
<tr>
<td>56.8</td>
<td>60.1</td>
<td>52.9</td>
<td>60.5</td>
<td>55.6</td>
<td>38.1</td>
</tr>
<tr>
<td>76.4</td>
<td>46.8</td>
<td>19.9</td>
<td>27.3</td>
<td>77.4</td>
<td>58.1</td>
</tr>
<tr>
<td>32.1</td>
<td>54.9</td>
<td>32.7</td>
<td>40.1</td>
<td>52.7</td>
<td>32.5</td>
</tr>
<tr>
<td>35.3</td>
<td>39.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Construct a frequency distribution starting with 10 as the lowest class beginning point and use a class width of 10. Construct a histogram and a frequency polygon for this frequency distribution and observe the shape of a normal distribution. On the basis of your results from these graphs, what does a normal distribution look like?

2.22 Use the data from Problem 2.20.
   a. Construct a histogram and a frequency polygon.
   b. Construct an ogive.

2.23 In a medium-size southern city, 86 houses are for sale, each having about 2000 ft² of floor space. The asking prices vary. The frequency distribution shown contains the price categories for the 86 houses. Construct a histogram, a frequency polygon, and an ogive from these data.

<table>
<thead>
<tr>
<th>ASKING PRICE</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ 60,000–under $ 70,000</td>
<td>21</td>
</tr>
<tr>
<td>70,000–under 80,000</td>
<td>27</td>
</tr>
<tr>
<td>80,000–under 90,000</td>
<td>18</td>
</tr>
<tr>
<td>90,000–under 100,000</td>
<td>11</td>
</tr>
<tr>
<td>100,000–under 110,000</td>
<td>6</td>
</tr>
<tr>
<td>110,000–under 120,000</td>
<td>3</td>
</tr>
</tbody>
</table>

2.24 Good, relatively inexpensive prenatal care often can prevent a lifetime of expense owing to complications resulting from a baby's low birth weight. A survey of a random sample of 57 new mothers asked them to estimate how much they spent on prenatal care. The researcher tallied the results and presented them in the frequency distribution shown. Use these data to construct a histogram, a frequency polygon, and an ogive.

<table>
<thead>
<tr>
<th>AMOUNT SPENT ON PRENATAL CARE</th>
<th>FREQUENCY OF NEW MOTHERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ 0–under $100</td>
<td>3</td>
</tr>
<tr>
<td>100–under 200</td>
<td>6</td>
</tr>
<tr>
<td>200–under 300</td>
<td>12</td>
</tr>
<tr>
<td>300–under 400</td>
<td>19</td>
</tr>
<tr>
<td>400–under 500</td>
<td>11</td>
</tr>
<tr>
<td>500–under 600</td>
<td>6</td>
</tr>
</tbody>
</table>

2.25 A consumer group surveyed food prices at 87 stores on the East Coast. Among the food prices being measured was that of sugar. From the data collected, the group constructed the frequency distribution of the prices of 5 lbs
of Domino's sugar in the stores surveyed. Compute a histogram, a frequency polygon, and an ogive for the following data.

<table>
<thead>
<tr>
<th>PRICE</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.75–under $1.90</td>
<td>9</td>
</tr>
<tr>
<td>1.90–under 2.05</td>
<td>14</td>
</tr>
<tr>
<td>2.05–under 2.20</td>
<td>17</td>
</tr>
<tr>
<td>2.20–under 2.35</td>
<td>16</td>
</tr>
<tr>
<td>2.35–under 2.50</td>
<td>18</td>
</tr>
<tr>
<td>2.50–under 2.65</td>
<td>8</td>
</tr>
<tr>
<td>2.65–under 2.80</td>
<td>5</td>
</tr>
</tbody>
</table>

2.26 The top music genres according to SoundScan for the year 1998 are R&B, Alternative (Rock) Music, Rap, and Country. These and other music genres along with the number of albums sold in each (in millions) are shown.

<table>
<thead>
<tr>
<th>GENRE</th>
<th>ALBUMS SOLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;B</td>
<td>146.4</td>
</tr>
<tr>
<td>Alternative</td>
<td>102.6</td>
</tr>
<tr>
<td>Rap</td>
<td>73.7</td>
</tr>
<tr>
<td>Country</td>
<td>64.5</td>
</tr>
<tr>
<td>Soundtrack</td>
<td>56.4</td>
</tr>
<tr>
<td>Metal</td>
<td>26.6</td>
</tr>
<tr>
<td>Classical</td>
<td>14.8</td>
</tr>
<tr>
<td>Latin</td>
<td>14.5</td>
</tr>
</tbody>
</table>

Construct a pie chart for these data displaying the percentage of the whole that each of these genre represents.

2.27 We show a list of the industries with the largest total release of toxic chemicals in 1996 according to the U.S. Environmental Protection Agency. Construct a pie chart to depict this information.

<table>
<thead>
<tr>
<th>INDUSTRY</th>
<th>TOTAL RELEASE (POUNDS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemicals</td>
<td>785,178,163</td>
</tr>
<tr>
<td>Primary metals</td>
<td>564,535,183</td>
</tr>
<tr>
<td>Paper</td>
<td>227,563,372</td>
</tr>
<tr>
<td>Plastics</td>
<td>116,409,291</td>
</tr>
<tr>
<td>Transportation</td>
<td>111,352,769</td>
</tr>
<tr>
<td>Fabricated metals</td>
<td>90,254,367</td>
</tr>
<tr>
<td>Food</td>
<td>83,303,395</td>
</tr>
<tr>
<td>Petroleum</td>
<td>68,887,258</td>
</tr>
<tr>
<td>Electrical</td>
<td>41,765,377</td>
</tr>
</tbody>
</table>

2.28 A research organization selected 50 U.S. towns with 1990 populations of between 4000 and 6000 as a sample to represent small towns for survey purposes. The populations of these towns follow.

<table>
<thead>
<tr>
<th>POPULATION</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>4420</td>
<td>5221</td>
</tr>
<tr>
<td>5049</td>
<td>4299</td>
</tr>
<tr>
<td>5338</td>
<td>5831</td>
</tr>
<tr>
<td>4653</td>
<td>5737</td>
</tr>
<tr>
<td>4730</td>
<td>4512</td>
</tr>
<tr>
<td>4758</td>
<td>5090</td>
</tr>
<tr>
<td>4866</td>
<td>5431</td>
</tr>
<tr>
<td>4216</td>
<td>5923</td>
</tr>
</tbody>
</table>

Construct a stem and leaf plot for the data, letting each leaf contain two digits.

2.29 Listed here are 30 different weekly Dow Jones industrial stock averages.

<table>
<thead>
<tr>
<th>STOCK AVERAGE</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2656</td>
<td>2301</td>
</tr>
<tr>
<td>2742</td>
<td>2830</td>
</tr>
<tr>
<td>2200</td>
<td>2764</td>
</tr>
<tr>
<td>2976</td>
<td>2375</td>
</tr>
<tr>
<td>2344</td>
<td>2760</td>
</tr>
<tr>
<td>2996</td>
<td>2437</td>
</tr>
</tbody>
</table>

2.30 The U.S. Department of Transportation keeps track of the percentage of flights arriving within 15 minutes of their scheduled arrivals. The following data are for 1990.

<table>
<thead>
<tr>
<th>AIRLINE</th>
<th>ON-TIME PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan Am</td>
<td>83.8</td>
</tr>
<tr>
<td>America West</td>
<td>83.8</td>
</tr>
<tr>
<td>Northwest</td>
<td>82.1</td>
</tr>
<tr>
<td>Eastern</td>
<td>81.2</td>
</tr>
<tr>
<td>USAir</td>
<td>80.8</td>
</tr>
<tr>
<td>Southwest</td>
<td>80.8</td>
</tr>
</tbody>
</table>

Use this information to construct a stem and leaf plot. Let the leaf be the tenths of a percent.

INTERPRETING THE OUTPUT

2.31 Suppose 150 shoppers at an upscale mall are interviewed and one of the questions asked is the household income. Study the MINITAB histogram of these data shown below and discuss what can be learned about the shoppers.

![Histogram of Household Income of Mall Shoppers](image)

2.32 We show an Excel-produced pie chart representing physician specialties. What does the chart tell you about the various specialties?
2.33 Suppose 100 CPA firms are surveyed to determine how many audits they perform over a certain time. The data are summarized using the MINITAB stem and leaf plot shown. What can you learn about the number of audits being performed by these firms from this plot?

**CHARACTER STEM AND LEAF DISPLAY**

<table>
<thead>
<tr>
<th>Stem and leaf of Number of Audits</th>
<th>N = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf Unit</td>
<td>1.0</td>
</tr>
<tr>
<td>9</td>
<td>1 222333333</td>
</tr>
<tr>
<td>16</td>
<td>1 4445555</td>
</tr>
<tr>
<td>26</td>
<td>1 6666667777</td>
</tr>
<tr>
<td>35</td>
<td>1 88889999</td>
</tr>
<tr>
<td>39</td>
<td>2 0001</td>
</tr>
<tr>
<td>44</td>
<td>2 22333</td>
</tr>
<tr>
<td>49</td>
<td>2 55555</td>
</tr>
<tr>
<td>(9)</td>
<td>2 6777777777</td>
</tr>
<tr>
<td>42</td>
<td>2 8888899</td>
</tr>
<tr>
<td>35</td>
<td>3 000111</td>
</tr>
<tr>
<td>29</td>
<td>3 223333</td>
</tr>
<tr>
<td>23</td>
<td>3 44455555</td>
</tr>
<tr>
<td>15</td>
<td>3 67777</td>
</tr>
<tr>
<td>10</td>
<td>3 889</td>
</tr>
<tr>
<td>7</td>
<td>4 0011</td>
</tr>
<tr>
<td>3</td>
<td>4 222</td>
</tr>
</tbody>
</table>
CALCULATING THE STATISTICS

3.41 The 2000 U.S. Census asks every household to report information on each person living there. Suppose a sample of 30 households is selected and the number of persons living in each is reported as follows:

2 3 1 2 6 4 2 1 5 3 2 3 1 2 2 1 3 1 2 2 4 2 1 2 8 3 2 1 1 3

Compute the mean, median, mode, range, lower and upper quartiles, and interquartile range for these data.

3.42 The 2000 U.S. Census also asks for each person’s age. Suppose that a sample of 40 households is taken from the census data and the age of the first person recorded on the census form is given as follows.

42 29 31 38 55 27 28
33 49 70 25 21 38 47
63 22 38 52 50 41 19
22 29 81 52 26 35 38
29 31 48 26 33 42 58
40 32 24 34 25

Compute $P_{10}$, $P_{80}$, $Q_1$, $Q_3$, the interquartile range, and the range for these data.

3.43 According to the National Association of Investment Clubs, PepsiCo, Inc., is the most popular stock with investment clubs with 11,388 clubs holding PepsiCo stock. The Intel Corp. is a close second, followed by Motorola, Inc. We show a list of the most popular stocks with investment clubs. Compute the mean, median, $P_{50}$, $P_{60}$, $P_{90}$, $Q_1$, $Q_3$, range, and interquartile range for these figures.

<table>
<thead>
<tr>
<th>COMPANY</th>
<th>NUMBER OF CLUBS HOLDING STOCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>PepsiCo, Inc.</td>
<td>11388</td>
</tr>
<tr>
<td>Intel Corp.</td>
<td>11019</td>
</tr>
<tr>
<td>Motorola, Inc.</td>
<td>9863</td>
</tr>
<tr>
<td>Tricon Global Restaurants</td>
<td>9168</td>
</tr>
<tr>
<td>Merck &amp; Co., Inc.</td>
<td>8687</td>
</tr>
<tr>
<td>AFLAC Inc.</td>
<td>6796</td>
</tr>
<tr>
<td>Diebold, Inc.</td>
<td>6552</td>
</tr>
<tr>
<td>McDonald’s Corp.</td>
<td>6498</td>
</tr>
<tr>
<td>Coca-Cola Co.</td>
<td>6101</td>
</tr>
<tr>
<td>Lucent Technologies</td>
<td>5563</td>
</tr>
<tr>
<td>Home Depot, Inc.</td>
<td>5414</td>
</tr>
<tr>
<td>Clayton Homes, Inc.</td>
<td>5390</td>
</tr>
<tr>
<td>RPM, Inc.</td>
<td>5033</td>
</tr>
<tr>
<td>Cisco Systems, Inc.</td>
<td>4541</td>
</tr>
<tr>
<td>General Electric Co.</td>
<td>4507</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>4464</td>
</tr>
<tr>
<td>Microsoft Corp.</td>
<td>4152</td>
</tr>
<tr>
<td>Wendy’s International, Inc.</td>
<td>4150</td>
</tr>
<tr>
<td>Walt Disney Co.</td>
<td>3999</td>
</tr>
<tr>
<td>AT&amp;T Corp.</td>
<td>3619</td>
</tr>
</tbody>
</table>

3.44 *Editor & Publisher International Yearbook* published a listing of the top 10 daily newspapers in the United States, as shown here. Use these population data to compute a mean and a standard deviation. The figures are given in average daily circulation from Monday through Friday. Because the numbers are large, it may save you some effort to recode the data. One way to recode these data is to move the decimal point six places to the left (e.g., 1,774,880 becomes 1.77488). If you recode the data this way, the resulting mean and standard deviation will be correct for the recoded data. To rewrite the answers so that they are correct for the original data, move the decimal point back to the right six places in the answers.

<table>
<thead>
<tr>
<th>NEWSPAPER</th>
<th>AVERAGE DAILY CIRCULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wall Street Journal</td>
<td>1,774,880</td>
</tr>
<tr>
<td>USA Today</td>
<td>1,629,665</td>
</tr>
<tr>
<td>New York Times</td>
<td>1,074,741</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>1,050,176</td>
</tr>
<tr>
<td>Washington Post</td>
<td>775,894</td>
</tr>
<tr>
<td>(N.Y.) Daily News</td>
<td>721,256</td>
</tr>
<tr>
<td>Chicago Tribune</td>
<td>653,554</td>
</tr>
<tr>
<td>Newday</td>
<td>568,914</td>
</tr>
<tr>
<td>Houston Chronicle</td>
<td>549,101</td>
</tr>
<tr>
<td>Chicago Sun-Times</td>
<td>484,379</td>
</tr>
</tbody>
</table>

3.45 We show the companies with the largest oil refining capacity in the world according to the *Petroleum Intelligence Weekly*. Use these population data and answer the questions.

<table>
<thead>
<tr>
<th>COMPANY</th>
<th>CAPACITY (1000's BARRELS PER DAY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exxon</td>
<td>4273</td>
</tr>
<tr>
<td>Royal Dutch/Shell</td>
<td>3791</td>
</tr>
<tr>
<td>China Petrochemical Corp.</td>
<td>2867</td>
</tr>
<tr>
<td>Petróleos de Venezuela</td>
<td>2437</td>
</tr>
<tr>
<td>Mobil</td>
<td>2297</td>
</tr>
<tr>
<td>Saudi Arabian Oil Co.</td>
<td>1970</td>
</tr>
<tr>
<td>British Petroleum</td>
<td>1965</td>
</tr>
<tr>
<td>Chevron</td>
<td>1661</td>
</tr>
<tr>
<td>Petrobras</td>
<td>1540</td>
</tr>
<tr>
<td>Texaco</td>
<td>1532</td>
</tr>
<tr>
<td>Petróleos Mexicanos (Pemex)</td>
<td>1520</td>
</tr>
<tr>
<td>National Iranian Oil Co.</td>
<td>1092</td>
</tr>
</tbody>
</table>

a. What are the values of the mean and the median? Compare the answers and state which you prefer as a measure of location for these data and why.

b. What are the values of the range and interquartile range? How do they differ?

c. What are the values of variance and standard deviation for these data?

d. What is the $Z$ score for Texaco? What is the $Z$ score for Mobil? Interpret these $Z$ scores.
3.46 The U.S. Department of the Interior’s Bureau of Mines releases figures on mineral production. Following are the 10 leading states in nonfuel mineral production in terms of the percentage of the U.S. total.

<table>
<thead>
<tr>
<th>STATE</th>
<th>PERCENT OF U.S. TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>8.91</td>
</tr>
<tr>
<td>Nevada</td>
<td>7.69</td>
</tr>
<tr>
<td>California</td>
<td>7.13</td>
</tr>
<tr>
<td>Georgia</td>
<td>4.49</td>
</tr>
<tr>
<td>Utah</td>
<td>4.46</td>
</tr>
<tr>
<td>Florida</td>
<td>4.42</td>
</tr>
<tr>
<td>Texas</td>
<td>4.31</td>
</tr>
<tr>
<td>Minnesota</td>
<td>4.06</td>
</tr>
<tr>
<td>Michigan</td>
<td>3.96</td>
</tr>
<tr>
<td>Missouri</td>
<td>3.34</td>
</tr>
</tbody>
</table>


gave responses ranging from 1 to 100. The agency’s analyst organizes the figures into a frequency distribution.

<table>
<thead>
<tr>
<th>NUMBER OF EMPLOYEES WORKING IN TELEMARKETING</th>
<th>NUMBER OF COMPANIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–under 20</td>
<td>32</td>
</tr>
<tr>
<td>20–under 40</td>
<td>16</td>
</tr>
<tr>
<td>40–under 60</td>
<td>13</td>
</tr>
<tr>
<td>60–under 80</td>
<td>10</td>
</tr>
<tr>
<td>80–under 100</td>
<td>19</td>
</tr>
</tbody>
</table>

a. Compute the mean, median, and mode for this distribution.

b. Compute the standard deviation for these data.

3.47 The radio music listener market is diverse. Listener formats might include adult contemporary, album rock, top 40, oldies, rap, country and western, classical, and jazz. In targeting audiences, market researchers need to be concerned about the ages of the listeners attracted to particular formats. Suppose a market researcher surveyed a sample of 170 listeners of oldies stations and obtained the following age distribution.

<table>
<thead>
<tr>
<th>AGE</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>15–under 20</td>
<td>9</td>
</tr>
<tr>
<td>20–under 25</td>
<td>16</td>
</tr>
<tr>
<td>25–under 30</td>
<td>27</td>
</tr>
<tr>
<td>30–under 35</td>
<td>44</td>
</tr>
<tr>
<td>35–under 40</td>
<td>42</td>
</tr>
<tr>
<td>40–under 45</td>
<td>23</td>
</tr>
<tr>
<td>45–under 50</td>
<td>7</td>
</tr>
<tr>
<td>50–under 55</td>
<td>2</td>
</tr>
</tbody>
</table>

a. What are the mean, median, and modal ages of oldies listeners?

b. What are the variance and standard deviation of the ages of oldies listeners?

3.48 A research agency administers a demographic survey to 90 telemarketing companies to determine the size of their operations. When asked to report how many employees now work in their telemarketing operation, the companies
3.52 During the 1990s, businesses were expected to show a lot of interest in Central and Eastern European countries. As new markets begin to open, American business people need to gain a better understanding of the market potential there. The following are the per capita GNP figures for eight of these European countries published by the World Almanac.

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>PER CAPITA INCOME (U.S. $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>1290</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>4630</td>
</tr>
<tr>
<td>Croatia</td>
<td>4300</td>
</tr>
<tr>
<td>Germany</td>
<td>20400</td>
</tr>
<tr>
<td>Hungary</td>
<td>7500</td>
</tr>
<tr>
<td>Poland</td>
<td>6400</td>
</tr>
<tr>
<td>Romania</td>
<td>5200</td>
</tr>
<tr>
<td>Bosnia/Herzegovina</td>
<td>600</td>
</tr>
</tbody>
</table>

a. Compute the mean and standard deviation for Albania, Bulgaria, Croatia, and Germany.
b. Compute the mean and standard deviation for Hungary, Poland, Romania, and Bosnia/Herzegovina.
c. Use a coefficient of variation to compare the two standard deviations. Treat the data as population data.

3.53 According to the Bureau of Labor Statistics, the average annual salary of a worker in Detroit, Michigan, is $35,748. Suppose the median annual salary for a worker in this group is $31,369 and the mode is $29,500. Is the distribution of salaries for this group skewed? If so, how and why? Which of these measures of central tendency would you use to describe these data? Why?

3.54 According to the U.S. Army Corps of Engineers, the top 20 U.S. ports, ranked by total tonnage (in million tons), were as follows.

<table>
<thead>
<tr>
<th>PORT</th>
<th>TOTAL TONNAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port of South Louisiana, LA</td>
<td>189.8</td>
</tr>
<tr>
<td>Houston, TX</td>
<td>148.2</td>
</tr>
<tr>
<td>New York, NY</td>
<td>131.6</td>
</tr>
<tr>
<td>New Orleans, LA</td>
<td>83.7</td>
</tr>
<tr>
<td>Baton Rouge, LA</td>
<td>81.0</td>
</tr>
<tr>
<td>Corpus Christi, TX</td>
<td>80.5</td>
</tr>
<tr>
<td>Valdez Harbor, AK</td>
<td>77.1</td>
</tr>
<tr>
<td>Port of Plaquemines, LA</td>
<td>66.9</td>
</tr>
<tr>
<td>Long Beach, CA</td>
<td>58.4</td>
</tr>
<tr>
<td>Texas City, TX</td>
<td>56.4</td>
</tr>
<tr>
<td>Mobile, AL</td>
<td>50.9</td>
</tr>
<tr>
<td>Pittsburgh, PA</td>
<td>50.9</td>
</tr>
<tr>
<td>Norfolk Harbor, VA</td>
<td>49.3</td>
</tr>
<tr>
<td>Tampa Harbor, FL</td>
<td>49.3</td>
</tr>
<tr>
<td>Lake Charles, LA</td>
<td>49.1</td>
</tr>
<tr>
<td>Los Angeles, CA</td>
<td>45.7</td>
</tr>
<tr>
<td>Baltimore Harbor, MD</td>
<td>43.6</td>
</tr>
<tr>
<td>Philadelphia, PA</td>
<td>41.9</td>
</tr>
<tr>
<td>Duluth-Superior, MN</td>
<td>41.4</td>
</tr>
<tr>
<td>Port Arthur, TX</td>
<td>37.2</td>
</tr>
</tbody>
</table>

a. Construct a box and whisker plot for these data.
b. Discuss the shape of the distribution from the plot.

c. Are there outliers?
d. What are they and why do you think they are outliers?

3.55 Runzheimer International publishes data on overseas business travel costs. They report that the average per diem total for a business traveler in Paris, France, is $349. Suppose the shape of the distribution of the per diem costs of a business traveler to Paris is unknown, but that 53% of the per diem figures are between $317 and $381. What is the value of the standard deviation? The average per diem total for a business traveler in Moscow is $415. If the shape of the distribution of per diem costs of a business traveler in Moscow is unknown and if 83% of the per diem costs in Moscow lie between $371 and $459, what is the standard deviation?

**INTERPRETING THE OUTPUT**

3.56 The American Banker has compiled a list of the top 100 banking companies in the world according to total assets. Leading the list is the Bank of Tokyo-Mitsubishi Ltd., followed by the Deutsche Bank AG. Below is an Excel descriptive statistics output for the variable total assets (in $ millions) for these 100 banks. Study the output and describe in your own words what you can learn about the assets of these top 100 world banks.

<table>
<thead>
<tr>
<th>Top World Banks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Sample variance</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Sum</td>
</tr>
<tr>
<td>Count</td>
</tr>
</tbody>
</table>

3.57 Hispanic Business, Inc., compiled a list of the top advertisers cultivating the Hispanic market. These data were entered into a MINITAB spreadsheet and analyzed using the graphical descriptive statistics feature. The dollar figures are in million dollar units. Study the output and describe the expenditures of these top Hispanic market advertisers.

| VARIABLE: MEDIA EXPEND |
|------------------------|------------------|
| Anderson-Darling Normality Test | 4.323 |
| P-Squared:               | 0.000 |
| Mean                    | 7.8560 |
| Standard deviation      | 5.8860 |
| Variance                | 34.6455 |
| Skewness                | 3.6214 |
| Kurtosis                | 17.7851 |
| N                       | 50    |
3.58 There are many large companies located around the world. The number of employees for 46 of the largest employers with headquarters outside the United States were analyzed with Excel’s descriptive statistics feature. The data follow. Summarize what you have learned about the number of employees for these companies by studying this output.

<table>
<thead>
<tr>
<th>Large Employers Outside of the United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard error</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Sample variance</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Sum</td>
</tr>
<tr>
<td>Count</td>
</tr>
</tbody>
</table>

3.59 The Competitive Media Reporting and Publishers Information Bureau compiled a list of the top 25 advertisers in the United States for a recent year. The total advertising expenditures for each company ($1000s) were analyzed using Minitab’s numerical descriptive statistics feature and its boxplot feature, both of which are displayed. Study this output and summarize the expenditures of the top 25 advertisers in your own words.

<table>
<thead>
<tr>
<th>DESCRIPTIVE STATISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Top 25 A</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>TnMean</td>
</tr>
<tr>
<td>StDev</td>
</tr>
<tr>
<td>SE Mean</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Q₁</td>
</tr>
<tr>
<td>Q₃</td>
</tr>
</tbody>
</table>

Top Twenty-five Advertisers

---

The diagram shows a boxplot for the total advertising expenditures of the top 25 advertisers, with the Y-axis ranges from 400,000 to 2,400,000.
4.34 Use the values in the contingency table to solve the equations given.

\[
\begin{array}{c|cc}
 & D & E \\
\hline
A & 10 & 20 \\
B & 15 & 5 \\
C & 30 & 15 \\
\end{array}
\]

Variable 2

a. \( P(E) = \) 

b. \( P(B \cup D) = \) 

c. \( P(A \cap E) = \) 

d. \( P(B \mid E) = \) 

e. \( P(A \cup B) = \) 

f. \( P(B \cap C) = \) 

g. \( P(D \mid C) = \) 

h. \( P(A \mid B) = \) 

i. Are variables 1 and 2 independent? Why or why not?

4.35 Use the values in the contingency table to solve the equations given.

\[
\begin{array}{c|cccc}
 & D & E & F & G \\
\hline
A & 3 & 9 & 7 & 12 \\
B & 8 & 4 & 6 & 4 \\
C & 10 & 5 & 3 & 7 \\
\end{array}
\]

a. \( P(F \cap A) = \) 

b. \( P(A \mid B) = \) 

c. \( P(B) = \) 

d. \( P(E \cap F) = \) 

e. \( P(D \mid B) = \) 

f. \( P(B \mid D) = \) 

g. \( P(D \cup C) = \) 

h. \( P(F) = \)
The following probability matrix contains a breakdown on the age and gender of U.S. physicians in a recent year, as reported by the American Medical Association.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>&lt;35</th>
<th>35–44</th>
<th>45–54</th>
<th>55–64</th>
<th>&gt;65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>.11</td>
<td>.20</td>
<td>.19</td>
<td>.12</td>
<td>.16</td>
<td>.78</td>
</tr>
<tr>
<td>Female</td>
<td>.07</td>
<td>.08</td>
<td>.04</td>
<td>.02</td>
<td>.01</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.18</td>
<td>.28</td>
<td>.23</td>
<td>.14</td>
<td>.17</td>
</tr>
</tbody>
</table>

a. What is the probability that one randomly selected physician is 35–44 years old?
b. What is the probability that one randomly selected physician is both a woman and 45–54 years old?
c. What is the probability that one randomly selected physician is a man or is 35–44 years old?
d. What is the probability that one randomly selected physician is less than 35 years old or 55–64 years old?
e. What is the probability that one randomly selected physician is a woman if she is 45–54 years old?
f. What is the probability that a randomly selected physician is neither a woman nor 55–64 years old?

**TESTING YOUR UNDERSTANDING**

4.37 **Purchasing Survey** asked purchasing professionals what sales traits impressed them the most in a sales representative. Seventy-eight percent selected “thoroughness.” Forty percent responded “knowledge of your own product.” The purchasing professionals were allowed to list more than one trait. Suppose 27% of the purchasing professionals listed both “thoroughness” and “knowledge of your own product” as sales traits that impressed them the most. A purchasing professional is randomly sampled.

a. What is the probability that the professional selected “thoroughness” or “knowledge of your own product”?
b. What is the probability that the professional selected neither “thoroughness” nor “knowledge of your own product”?
c. If it is known that the professional selected “thoroughness,” what is the probability that the professional selected “knowledge of your own product”?
d. What is the probability that the professional did not select “thoroughness” and did select “knowledge of your own product”?

4.38 The U.S. Bureau of Labor Statistics publishes data on the benefits offered by small companies to their employees. Only 42% offer retirement plans while 61% offer life insurance. Suppose 33% offer both retirement plans and life insurance as benefits. If a small company is randomly selected, what is the probability that:

a. The company offers a retirement plan given that they offer life insurance?
b. The company offers life insurance given that they offer a retirement plan?
c. The company offers life insurance or a retirement plan?
d. The company offers a retirement plan and does not offer life insurance?
e. The company does not offer life insurance if it is known that they offer a retirement plan?

4.39 According to Link Resources, 16% of the U.S. population is technology-driven. However, these figures vary by region. For example, in the West the figure is 20% and in the Northeast the figure is 17%. Twenty-one percent of the U.S. population in general is in the West and 20% of the U.S. population is in the Northeast. Suppose an American is chosen randomly.

a. What is the probability that the person lives in the West and is a technology-driven person?
b. What is the probability that the person lives in the Northeast and is a technology-driven person?
c. Suppose the chosen person is known to be technology-driven. What is the probability that the person lives in the West?
d. Suppose the chosen person is known not to be technology-driven. What is the probability that the person lives in the Northeast?
e. Suppose the chosen person is known to be technology-driven. What is the probability that the person lives in neither the West nor the Northeast?

4.40 In a certain city, 30% of the families have a MasterCard, 20% have an American Express card, and 25% have a Visa card. Eight percent of the families have both a MasterCard and an American Express card. Twelve percent have both a Visa card and a MasterCard. Six percent have both an American Express card and a Visa card.

a. What is the probability that a family that has either a Visa card or an American Express card?
b. If a family has a MasterCard, what is the probability that it has a Visa card?
c. If a family has a Visa card, what is the probability that it has a MasterCard?
d. Is possession of a Visa card independent of possession of a MasterCard? Why or why not?
e. Is possession of an American Express card mutually exclusive of possession of a Visa card?

4.41 A few years ago, a survey commissioned by *The World Almanac* and Maturity News Service reported that 51% of the respondents did not believe the Social Security system will be secure in 20 years. Of the respondents who were age 45 or older, 70% believed the system will be
secure in 20 years. Of the people surveyed, 57% were under age 45. One respondent is selected randomly.

a. What is the probability that the person is age 45 or older?
b. What is the probability that the person is younger than age 45 and believes that the Social Security system will be secure in 20 years?
c. If the person selected believes the Social Security system will be secure in 20 years, what is the probability that the person is 45 years old or older?
d. What is the probability that the person is younger than age 45 or believes the Social Security system will not be secure in 20 years?

A telephone survey conducted by the Maritz Marketing Research company found that 43% of Americans expect to save more money next year than they saved last year. Forty-five percent of those surveyed plan to reduce debt next year. Of those who expect to save more money next year, 81% plan to reduce debt next year. An American is selected randomly.

a. What is the probability that this person expects to save more money next year and plans to reduce debt next year?
b. What is the probability that this person expects to save more money next year or plans to reduce debt next year?
c. What is the probability that this person neither expects to save more money next year nor plans to reduce debt next year?
d. What is the probability that this person expects to save more money next year and does not plan to reduce debt next year?

The Steelcase Workplace Index studied the types of work-related activities that Americans did while on vacation in the summer. Among other things, 40% read work-related material. Thirty-four percent checked in with the boss. Respondents to the study were allowed to select more than one activity. Suppose that of those who read work-related material, 78% checked in with the boss. One of these survey respondents is selected randomly. What is the probability that while on vacation this respondent:

a. Checked in with the boss and read work-related material?
b. Neither read work-related material nor checked in with the boss?
c. Read work-related material given that he checked in with the boss?
d. Did not check in with the boss given that he read work-related material?
e. Did not check in with the boss given that he did not read work-related material?
f. Construct a probability matrix for this problem.

Health Rights Hotline published the results of a survey of 2400 people in Northern California in which consumers were asked to share their complaints about managed care. The number one complaint was denial of care, with 17% of the participating consumers selecting it. Several other complaints were noted including inappropriate care (14%), customer service (14%), payment disputes (11%), specialty care (10%), delays in getting care (8%), and prescription drugs (7%). These complaint categories are mutually exclusive. Assume that the results of this survey can be inferred to all managed care consumers. If a managed care consumer is randomly selected, what is the probability that:

a. The consumer complains about payment disputes or specialty care?
b. The consumer complains about prescription drugs and customer service?
c. The consumer complains about inappropriate care given that she complains about specialty care?
d. The consumer does not complain about delays in getting care nor does she complain about payment disputes?

Companies use employee training for various reasons including employee loyalty, certification, quality, and process improvement. In a national survey of companies, BI Learning Systems reported that 56% percent of the responding companies named employee retention as a top reason for training. Suppose 36% of the companies replied that they use training for process improvement and for employee retention. In addition, suppose that of the companies that use training for process improvement, 90% use training for employee retention. A company that uses training is randomly selected.

a. What is the probability that the company uses training for employee retention and not for process improvement?
b. If it is known that the company uses training for employee retention, what is the probability that it uses training for process improvement?
c. What is the probability that the company uses training for process improvement?
d. What is the probability that the company uses training for employee retention or process improvement?
e. What is the probability that the company neither uses training for employee retention nor uses training for process improvement?
f. Suppose it is known that the company does not use training for process improvement. What is the probability that the company does use training for employee retention?

Pitney Bowes surveyed 302 directors and vice presidents of marketing at large and midsize U.S. companies to determine what they believe is the best vehicle for educating decision makers on complex issues in selling products and services. The highest percentage of companies chose direct mail/catalogs, followed by direct sales/sales rep. Direct mail/catalogs was selected by 38% of the com-
companies. None of the companies selected both direct mail/catalogs and direct sales/sales rep. Suppose also that 41% selected neither direct mail/catalogs nor direct sales/sales rep. If one of these companies is selected randomly and their top marketing person interviewed about this matter, what is the probability that she:

a. Selected direct mail/catalogs and did not select direct sales/sales rep?

b. Selected direct sales/sales rep?

c. Selected direct sales/sales rep given that she selected direct mail/catalogs?

d. Did not select direct mail/catalogs given that she did not select direct sales/sales rep?

4.47 A small independent physicians' practice has three doctors. Doctor Sarabia sees 41% of the patients, Doctor Tran sees 32%, and Doctor Jackson sees the rest. Doctor Sarabia requests blood tests on 5% of her patients, Doctor Tran requests blood tests on 8% of his patients, and Doctor Jackson requests blood tests on 6% of her patients. An auditor randomly selects a patient from the past week and discovers that the patient has had a blood test as a result of his physician visit. Knowing this, what is the probability that the patient saw Doctor Sarabia? For what percentage of all patients at this practice are blood tests requested?

4.48 A survey by the Arthur Anderson Enterprise Group/National Small Business United attempted to determine what the leading challenges are for the growth and survival of small businesses. While the economy and finding qualified workers were the leading challenges, there were several others listed in the results of the study. Among those are regulations, listed by 30% of the companies, and the tax burden, listed by 35%. Suppose that 71% of the companies listing regulations as a challenge listed the tax burden as a challenge. Assume these percentages hold for all small businesses. If a small business is randomly selected, what is the probability that it:

a. Lists both the tax burden and regulations as a challenge?

b. Lists either the tax burden or regulations as a challenge?

c. Lists either the tax burden or regulations but not both as a challenge?

d. Lists regulations as a challenge given that it lists the tax burden as a challenge?

e. Does not list regulations as a challenge given that it lists the tax burden as a challenge?

f. Does not list regulations as a challenge given that it does not list the tax burden as a challenge?

4.49 According to the Public Voice for Food and Health Policy, approximately 27% of all soup products in a recent year did not carry nutritional labeling. Approximately 83% of breakfast meats and about 59% of hot dog products did not have nutritional labeling. Assume that if these three groups of foods were combined, 60% would be soup products, 35% would be breakfast meats, and 5% would be hot dogs. A researcher is blindly given a food product from one of these three groups and is told that the product does have nutritional labeling. Revise the probabilities that the product is a soup product, a breakfast meat, and a hot dog product.

4.50 A survey conducted for Lifetime's daily half-hour series "The Great American TV Poll" asked Americans what they consider to be the most important thing in their lives. Twenty-nine percent said "good health," 21% responded "a happy marriage," and 40% replied "faith in God." Because they were asked which of these things is the most important thing, a respondent could not select more than one answer.

a. What is the probability that a person replied "a happy marriage," or "faith in God"?

b. What is the probability that a person replied "a happy marriage" or "faith in God" or "good health"?

c. What is the probability that a person replied "faith in God" and "good health"?

d. What is the probability that a person replied neither "faith in God" nor "good health" nor "a happy marriage"?
CALCULATING THE STATISTICS

5.34 Solve for the probabilities of the following binomial distribution problems by using the binomial formula.
   a. If \( n = 11 \) and \( p = .23 \), what is the probability that \( X = 4 \)?
   b. If \( n = 6 \) and \( p = .50 \), what is the probability that \( X \geq 1 \)?
   c. If \( n = 9 \) and \( p = .85 \), what is the probability that \( X > 7 \)?
   d. If \( n = 14 \) and \( p = .70 \), what is the probability that \( X \leq 3 \)?

5.35 Use Table A.2, Appendix A, to find the values of the following binomial distribution problems.
   a. \( P(X = 14|n = 20 \text{ and } p = .60) \)
   b. \( P(X < 5|n = 10 \text{ and } p = .30) \)
   c. \( P(X \geq 12|n = 15 \text{ and } p = .60) \)
   d. \( P(X > 20|n = 25 \text{ and } p = .40) \)

5.36 Use the Poisson formula to solve for the probabilities of the following Poisson distribution problems.
   a. If \( \lambda = 1.25 \), what is the probability that \( X = 4 \)?
   b. If \( \lambda = 6.37 \), what is the probability that \( X \leq 1 \)?
   c. If \( \lambda = 2.4 \), what is the probability that \( X > 5 \)?

5.37 Use Table A.3, Appendix A, to find the following Poisson distribution values.
   a. \( P(X = 3|\lambda = 1.8) \)
   b. \( P(X < 5|\lambda = 3.3) \)
   c. \( P(X \geq 3|\lambda = 2.1) \)
   d. \( P(2 < X \leq 5|\lambda = 4.2) \)

5.38 Solve the following problems by using the hypergeometric formula.
   a. If \( N = 6 \), \( n = 4 \), and \( A = 5 \), what is the probability that \( X = 3 \)?
   b. If \( N = 10 \), \( n = 3 \), and \( A = 5 \), what is the probability that \( X \leq 1 \)?
   c. If \( N = 13 \), \( n = 5 \), and \( A = 3 \), what is the probability that \( X \geq 2 \)?

TESTING YOUR UNDERSTANDING

5.39 In a study by Peter D. Hart Research Associates for the Nasdaq Stock Market, it was determined that 20% of all stock investors are retired people. In addition, 40% of all U.S. adults have invested in mutual funds. Suppose a random sample of 25 stock investors is taken. What is the probability that exactly seven are retired people? What is the probability that 10 or more are retired people? How many retired people would you expect to find in a random sample of 25 stock investors? Suppose a random sample of 20 U.S. adults is taken. What is the probability that exactly eight adults have invested in mutual funds? What is the probability that fewer than six adults have invested in mutual funds? What is the probability that none of the adults have invested in mutual funds? What is the probability that 12 or more adults have invested in mutual funds? For which exact number of adults is the probability the highest? How does this compare to the expected number?

5.40 A service station has a pump that distributes diesel fuel to automobiles. The station owner estimates that only about 3.2 cars use the diesel pump every 2 hours. Assume the arrivals of diesel pump users are Poisson distributed.
   a. What is the probability that three cars will arrive to use the diesel pump during a 1-hour period?
   b. Suppose the owner needs to shut down the diesel pump due to high costs. However, the owner hates to lose any business. What is the probability that no cars will arrive to use the diesel pump during a half-hour period?
   c. Suppose five cars arrive during a 1-hour period to use the diesel pump. What is the probability of five or more cars arriving during a 1-hour period to use the diesel pump? If this outcome actually occurred, what might you conclude?

5.41 In a particular manufacturing plant, two machines (A and B) produce a particular part. One machine (B) is newer and faster. In one 5-minute period, a lot consisting of 32 parts is produced. Twenty-two are produced by machine B and the rest by machine A. Suppose an inspector randomly samples a dozen of the parts from this lot.
   a. What is the probability that exactly three parts were produced by machine A?
   b. What is the probability that half of the parts were produced by each machine?
   c. What is the probability that all of the parts were produced by machine B?
   d. What is the probability that seven, eight, or nine parts were produced by machine B?

5.42 Suppose that, for every lot of 100 computer chips a company produces, an average of 1.4 are defective. Another company buys many lots of these chips at a time, from which one lot is selected randomly and tested for defects. If the tested lot contains more than three defects, the buyer will reject all the lots sent in that batch. What is the probability that the buyer will accept the lots? Assume that the defects per lot are Poisson-distributed.

5.43 The National Center for Health Statistics reports that 25% of all Americans between the ages of 65 and 74 have a chronic heart condition. Suppose you live in a state where the environment is conducive to good health and low stress and you believe the conditions in your state promote healthy hearts. To investigate this theory, you conduct a random telephone survey of 20 persons 65 to 74 years of age in your state.
5.44 A survey conducted for the Northwestern National Life Insurance Company revealed that 70% of American workers say job stress caused frequent health problems. One in three said they expected to burn out in the job in the near future. Thirty-four percent said they thought seriously about quitting their job last year because of workplace stress. Fifty-three percent said they were required to work more than 40 hours a week very often or somewhat often.

a. Suppose a random sample of 10 American workers is selected. What is the probability that more than seven of them say job stress caused frequent health problems? What is the expected number of workers who say job stress caused frequent health problems?
b. Suppose a random sample of 15 American workers is selected. What is the expected number of these sampled workers who say they will burn out in the near future? What is the probability that none of the workers say they will burn out in the near future?
c. Suppose a sample of seven workers is selected randomly. What is the probability that all seven say they are asked very often or somewhat often to work more than 40 hours a week? If this outcome actually happened, what might you conclude?

5.46 According to a recent survey, the probability that a passenger files a complaint with the Department of Transportation about a particular U.S. airline is 0.00014. Suppose 100,000 passengers who have flown this particular airline are randomly contacted.

a. What is the probability that exactly five passengers have filed complaints?
b. What is the probability that none of the passengers have filed complaints?
c. What is the probability that more than six passengers have filed complaints?

5.47 A hair stylist has been in business one year. Sixty percent of his customers are walk-in business. If he randomly samples eight of the people from last week’s list of customers, what is the probability that three or fewer were walk-ins? If this outcome actually occurred, what would be some of the explanations for it?

5.48 According to the U.S. Bureau of the Census, about 20% of Idaho residents live in metropolitan areas. This percentage is the lowest of all 50 states. A catalog sales company in Georgia has just purchased a list of Idaho consumers. Its market analyst randomly selects 25 people from this list.

a. What is the probability that exactly eight people live in metropolitan areas?
b. What is the probability that the analyst would get more than 10 people in this sample who live in metropolitan areas?
c. Suppose the analyst got more than 10 people who live in metropolitan areas from the group of 25. What might she conclude about the company’s list of Idaho consumers? What might she conclude about the census figure?

5.49 Suppose that, for every family vacation trip by car of more than 2000 miles, an average of 0.6 flat tires occurs. Suppose also that the distribution of the number of flat tires per trip of more than 2000 miles is Poisson. What is the probability that a family will take a trip of more than 2000 miles and have no flat tires? What is the probability that the family will have three or more flat tires on such a trip? Suppose trips are independent and the value of lambda holds for all trips of more than 2000 miles. If a family takes two trips of more than 2000 miles during a summer, what is the probability that the family will have no flat tires on either trip?

5.50 Editor and Publisher Yearbook releases figures on the top newspapers in the United States. Shown here are the top 25 Sunday newspapers in the United States ranked according to circulation.
Suppose a researcher wants to sample a portion of these newspapers and compare the sizes of the business sections of the Sunday papers. She randomly samples eight of these newspapers.

a. What is the probability that the sample contains exactly one newspaper located in New York state?
b. What is the probability that half of the newspapers are ranked in the top ten by circulation?
c. What is the probability that none of the newspapers are located in California?
d. What is the probability that exactly three of the newspapers are located in states that begin with the letter 'M'?

An office in Albuquerque has 24 workers including management. Eight of the workers commute to work from the west side of the Rio Grande River. Suppose six of the office workers are randomly selected.

a. What is the probability that all six workers commute from the west side of the Rio Grande?
b. What is the probability that none of the workers commute from the west side of the Rio Grande?
c. Which probability from parts (a) and (b) was greatest? Why do you think this is?
d. What is the probability that half of the workers do not commute from the west side of the Rio Grande?

According to the U.S. Bureau of the Census, 20% of the workers in Atlanta use public transportation. If 25 Atlanta workers are randomly selected, what is the expected number who use public transportation? Graph the binomial distribution for this sample. What is the mean and the standard deviation for this distribution? What is the probability that more than 12 of the selected workers use public transportation? Explain conceptually and from the graph why you would get this probability. Suppose you randomly sample 25 Atlanta workers and actually get 14 who use public transportation. Is this likely? How might you explain this result?

One of the earliest applications of the Poisson distribution was in analyzing incoming calls to a telephone switchboard. Analysts generally believe that random phone calls are Poisson distributed. Suppose phone calls to a switchboard arrive at an average rate of 2.4 calls per minute.

a. If an operator wants to take a 1-minute break, what is the probability that there will be no calls during a 1-minute interval?
b. If an operator can handle at most five calls per minute, what is the probability that the operator will be unable to handle the calls in any 1-minute period?
c. What is the probability that exactly three calls will arrive in a 2-minute interval?
d. What is the probability that one or fewer calls will arrive in a 15-second interval?

Only 1% of all American households do not have a color television set. A television marketing analyst randomly selects 160 American households.

a. How many households would you expect to not have a color television set?
b. What is the probability that eight or more households do not have a color television set?
c. What is the probability that between two and six households (inclusive) do not have a color television set?

Suppose that in the bookkeeping operation of a large corporation the probability of a recording error on any one billing is .005. Suppose the probability of a recording error from one billing to the next is constant and 1000 billings are randomly sampled by an auditor.

a. What is the probability that fewer than four billings contain a recording error?
b. What is the probability that more than 10 billings contain a billing error?
c. What is the probability that all 1000 billings contain no recording errors?

According to the American Medical Association, about 36% of all U.S. physicians under the age of 35 are women. Your company has just hired eight physicians under the age of 35 and none is a woman. If a group of women physicians under the age of 35 want to sue your company for discriminatory hiring practices, would they have a strong case based on these numbers? Use the binomial distribution to determine the probability of the company's hiring result occurring randomly and comment on the potential justification for a lawsuit.
The following table lists the 26 largest U.S. universities according to enrollment figures from the World Almanac.

<table>
<thead>
<tr>
<th>UNIVERSITY</th>
<th>ENROLLMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Texas at Austin</td>
<td>48,857</td>
</tr>
<tr>
<td>Ohio State University–Columbus</td>
<td>48,278</td>
</tr>
<tr>
<td>University of Minnesota</td>
<td>45,410</td>
</tr>
<tr>
<td>Arizona State University</td>
<td>44,255</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>42,603</td>
</tr>
<tr>
<td>University of Florida</td>
<td>41,713</td>
</tr>
<tr>
<td>University of Phoenix</td>
<td>41,467</td>
</tr>
<tr>
<td>Texas A&amp;M University–College Station</td>
<td>41,461</td>
</tr>
<tr>
<td>Pennsylvania State–University Park</td>
<td>40,538</td>
</tr>
<tr>
<td>University of Wisconsin–Madison</td>
<td>40,196</td>
</tr>
<tr>
<td>University of Michigan</td>
<td>36,995</td>
</tr>
<tr>
<td>New York University</td>
<td>36,684</td>
</tr>
<tr>
<td>University of Illinois–Champaign</td>
<td>36,019</td>
</tr>
<tr>
<td>Purdue University–West Lafayette, Indiana</td>
<td>35,715</td>
</tr>
<tr>
<td>University of California at Los Angeles</td>
<td>35,558</td>
</tr>
<tr>
<td>University of Washington</td>
<td>35,367</td>
</tr>
<tr>
<td>Indiana University–Bloomington</td>
<td>34,937</td>
</tr>
<tr>
<td>University of South Florida</td>
<td>34,036</td>
</tr>
<tr>
<td>University of Arizona</td>
<td>33,737</td>
</tr>
<tr>
<td>University of Maryland–College Park</td>
<td>32,711</td>
</tr>
<tr>
<td>Brigham Young University</td>
<td>32,161</td>
</tr>
<tr>
<td>University of Houston</td>
<td>31,602</td>
</tr>
<tr>
<td>Wayne State University (Detroit)</td>
<td>30,729</td>
</tr>
<tr>
<td>University of California at Berkeley</td>
<td>30,584</td>
</tr>
<tr>
<td>Florida State University</td>
<td>30,401</td>
</tr>
<tr>
<td>Florida International University</td>
<td>30,012</td>
</tr>
</tbody>
</table>

The Public Citizen’s Health Research Group studied the serious disciplinary actions that were taken during a recent year on nonfederal medical doctors in the United States. The national average was 3.84 serious actions per 1000 doctors. The state with the lowest number was Minnesota, with 1.6 serious actions per 1000 doctors. Assume that the number of serious actions per 1000 doctors in both the United States and in Minnesota are Poisson distributed.

a. What is the probability of randomly selecting 1000 U.S. doctors and finding out that there were no serious actions taken?

b. What is the probability of randomly selecting 2000 U.S. doctors and finding out that there were six serious actions taken?

c. What is the probability of randomly selecting 3000 Minnesota doctors and finding out that there were fewer than seven serious actions taken?

**INTERPRETING THE OUTPUT**

Study the MINITAB output. Discuss the type of distribution, the mean, standard deviation, and why the probabilities fall as they do.

**PROBABILITY DENSITY FUNCTION**

Binomial with n = 15 and p = 0.360000

<table>
<thead>
<tr>
<th>x</th>
<th>P(X = x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.0012</td>
</tr>
<tr>
<td>1.00</td>
<td>0.0104</td>
</tr>
<tr>
<td>2.00</td>
<td>0.0411</td>
</tr>
<tr>
<td>3.00</td>
<td>0.1002</td>
</tr>
<tr>
<td>4.00</td>
<td>0.1692</td>
</tr>
<tr>
<td>5.00</td>
<td>0.2093</td>
</tr>
<tr>
<td>6.00</td>
<td>0.1963</td>
</tr>
<tr>
<td>7.00</td>
<td>0.1419</td>
</tr>
<tr>
<td>8.00</td>
<td>0.0798</td>
</tr>
<tr>
<td>9.00</td>
<td>0.0349</td>
</tr>
<tr>
<td>10.00</td>
<td>0.0118</td>
</tr>
<tr>
<td>11.00</td>
<td>0.0030</td>
</tr>
<tr>
<td>12.00</td>
<td>0.0006</td>
</tr>
<tr>
<td>13.00</td>
<td>0.0001</td>
</tr>
<tr>
<td>14.00</td>
<td>0.0000</td>
</tr>
<tr>
<td>15.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

In one midwestern city, the government has 14 repossessed houses, which are evaluated to be worth about the same. Ten of the houses are on the north side of town and the rest are on the west side. A local contractor has submitted a bid in which he offered to purchase four of the houses. Which houses the contractor will get is subject to a random draw.

a. What is the probability that all four houses selected for the contractor will be on the north side of town?

b. What is the probability that all four houses selected for the contractor will be on the west side of town?

c. What is the probability that half of the houses selected for the contractor will be on the west side and half on the north side of town?

Study the Excel output. Explain the distribution in terms of shape and mean. Are these probabilities what you would expect? Why or why not?

<table>
<thead>
<tr>
<th>X VALUES</th>
<th>POISSON PROBABILITY: LAMBDa = 2.78</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0620</td>
</tr>
<tr>
<td>1</td>
<td>0.1725</td>
</tr>
<tr>
<td>2</td>
<td>0.2397</td>
</tr>
<tr>
<td>3</td>
<td>0.2221</td>
</tr>
<tr>
<td>4</td>
<td>0.1544</td>
</tr>
</tbody>
</table>

(continued)
5.62 Study the graphical output from Excel. Describe the distribution and explain why the graph takes the shape it does.

5.63 Study the MINITAB graph. Discuss the distribution including type, shape, and probability outcomes.

**ANALYZING THE DATABASES**

1. Use the manufacturing database. What is the probability that a randomly selected SIC code industry has a value of industry shipments equal to 2? Use this as the $p$ value for a binomial experiment. If you were to randomly select 12 SIC code industries, what is the probability that fewer than three would have a value of industry shipments equal to 2? If you were to randomly select 25 SIC code industries, what is the probability that exactly eight would have a value of industry shipments equal to 2?

2. Use the hospital database. In this population of 200, what is the breakdown between hospitals that are general medical hospitals and psychiatric hospitals? Using those figures as a breakdown of the population and the hypergeometric distribution, what is the probability of randomly selecting 16 hospitals from this database and getting exactly nine that are psychiatric hospitals? Using the number of hospitals in this database that are for-profit, compute $p = \text{probability that a hospital is for-profit}$. Now use the binomial formula to determine the probability of randomly selecting 30 hospitals and getting exactly 10 that are for-profit.

3. Use the financial database of chemical companies. If five of these companies are selected randomly, what is the probability that exactly three have a return on equity of 15% or more? *Hint: Use the hypergeometric distribution and a breakdown of this population of 19 companies to compute this probability. What is the probability of randomly selecting eight insurance companies and getting exactly four of them with average yields of less than 1%?*
CALCULATING THE STATISTICS

6.34 Data are uniformly distributed between the values of 6 and 14. Determine the value of \( f(x) \). What are the mean and standard deviation of this distribution? What is the probability of randomly selecting a value greater than 11? What is the probability of randomly selecting a value between 7 and 12?

6.35 Assume a normal distribution and find the following probabilities.
   a. \( P(X < 21|\mu = 25 \text{ and } \sigma = 4) \)
   b. \( P(X \geq 77|\mu = 50 \text{ and } \sigma = 9) \)
   c. \( P(X > 47|\mu = 50 \text{ and } \sigma = 6) \)
   d. \( P(13 < X < 29|\mu = 23 \text{ and } \sigma = 4) \)
   e. \( P(X \geq 105|\mu = 90 \text{ and } \sigma = 2.86) \)

6.36 Work the following binomial distribution problems by using the normal distribution. Check your answers by using Table A.2 to solve for the probabilities.
   a. \( P(X = 12|n = 25 \text{ and } p = .60) \)
   b. \( P(X > 5|n = 15 \text{ and } p = .50) \)
   c. \( P(X \leq 3|n = 10 \text{ and } p = .50) \)
   d. \( P(X \geq 8|n = 15 \text{ and } p = .40) \)

6.37 Find the probabilities for the following exponential distribution problems.
   a. \( P(X \geq 3|\lambda = 1.3) \)
   b. \( P(X < 2|\lambda = 2.0) \)
   c. \( P(1 \leq X \leq 3|\lambda = 1.65) \)
   d. \( P(X > 2|\lambda = .405) \)

TESTING YOUR UNDERSTANDING

6.38 The U.S. Bureau of Labor Statistics reports that of persons who usually work full time, the average number of hours worked per week is 43.4. Assume that the number of hours worked per week for those who usually work full time is normally distributed. Suppose 12% of these workers work more 48 hours. Based on this, what is the standard deviation of number of hours worked per week for these workers?

6.39 A U.S. Bureau of Labor Statistics survey showed that one in five people 16 years of age or older volunteers some of his or her time. If this figure holds for the entire population and if a random sample of 150 people 16 years of age or older is taken, what is the probability that more than 50 of those sampled do volunteer work?

6.40 An entrepreneur opened a small hardware store in a strip mall. During the first few weeks, business was slow, with the store averaging only one customer every 20 minutes in the morning. Assume that the random arrival of customers is Poisson distributed.
   a. What is the probability that at least 1 hour would elapse between customers?

6.41 In a recent year, the average price of a Microsoft Windows Upgrade was $90.28 according to PC Data. Assume that prices of the Microsoft Windows Upgrade that year were normally distributed, with a standard deviation of $8.53. If a retailer of computer software was randomly selected that year, what is the probability that the price of a Microsoft Windows Upgrade was below $80? What is the probability that the price was above $95? What is the probability that the price was between $83 and $87?

6.42 According to the U.S. Department of Agriculture, Alabama egg farmers produce millions of eggs every year. Suppose egg production per year in Alabama is normally distributed, with a standard deviation of 83 million eggs. If during only 3% of the years Alabama egg farmers produce more than 2655 million eggs, what is the mean egg production by Alabama farmers?

6.43 The U.S. Bureau of Labor Statistics releases figures on the number of full-time wage and salary workers with flexible schedules. The numbers of full-time wage and salary workers in each age category are almost uniformly distributed by age, with ages ranging from 18 to 65 years. If a worker with a flexible schedule is randomly drawn from the U.S. work force, what is the probability that he or she will be between 25 and 50 years of age? What is the mean value for this distribution? What is the height of the distribution?

6.44 A business convention holds its registration on Wednesday morning from 9:00 A.M. until 12:00 noon. Past history has shown that registrant arrivals follow a Poisson distribution at an average rate of 1.8 every 15 seconds. Fortunately, several facilities are available to register convention members.
   a. What is the average number of seconds between arrivals to the registration area for this conference based on past results?
   b. What is the probability that 25 seconds or more would pass between registration arrivals?
   c. What is the probability that less than 5 seconds will elapse between arrivals?
   d. Suppose the registration computers went down for a 1-minute period. Would this condition pose a problem? What is the probability that at least 1 minute will elapse between arrivals?

6.45 MPIF Research, Inc. lists the average monthly apartment rent in some of the most expensive apartment rental locations in the United States. According to their report, the average cost of renting an apartment in Minneapolis is
$951. Suppose that the standard deviation of the cost of renting an apartment in Minneapolis is $96 and that apartment rents in Minneapolis are normally distributed. If a Minneapolis apartment is randomly selected, what is the probability that the price is:

a. $1000 or more?
b. Between $900 and $1100?
c. Between $825 and $925?
d. Less than $700?

6.46 According to *The Wirthlin Report*, 24% of all workers say that their job is very stressful. If 60 workers are randomly selected, what is probability that 17 or more say that their job is very stressful? What is the probability that more than 22 say that their job is very stressful? What is the probability that between 8 and 12 (inclusive) say that their job is very stressful?

6.47 The U.S. Bureau of Labor Statistics reports that the average annual salary in the metropolitan Boston area is $34,383. Suppose annual salaries in the metropolitan Boston area are normally distributed, with a standard deviation of $4097. A Boston area worker is randomly selected.

a. What is the probability that the worker's annual salary is more than $40,000?
b. What is the probability that the worker's annual salary is less than $30,000?
c. What is the probability that the worker's annual salary is more than $20,000?
d. What is the probability that the worker's annual salary is between $27,000 and $36,000?

6.48 Suppose interarrival times at a hospital emergency room during a weekday are exponentially distributed, with an average interarrival time of 9 minutes. If the arrivals are Poisson distributed, what would the average number of arrivals per hour be? What is the probability that less than 5 minutes will elapse between any two arrivals?

6.49 Suppose the average speeds of passenger trains traveling from Newark, New Jersey, to Philadelphia, Pennsylvania, are normally distributed, with a mean average speed of 88 mph and a standard deviation of 6.4 mph.

a. What is the probability that a train will average less than 70 mph?
b. What is the probability that a train will average more than 80 mph?
c. What is the probability that a train will average between 90 and 100 mph?

6.50 The Conference Board published information on why companies expect to increase the number of part-time jobs and reduce full-time positions. Eighty-one percent of the companies said the reason was to get a flexible workforce. Suppose 200 companies that expect to increase the number of part-time jobs and reduce full-time positions are identified and contacted. What is the expected number of these companies that would agree that the reason is to get a flexible workforce? What is the probability that between 150 and 155 (not including the 150 or the 155) would give that reason? What is the probability that more than 158 would give that reason? What is the probability that fewer than 144 would give that reason?

6.51 According to the U.S. Bureau of the Census, about 75% of all commuters in the United States drive alone. Suppose 150 U.S. commuters are randomly sampled.

a. What is the probability that fewer than 105 commuters drive to work alone?
b. What is the probability that between 110 and 120 (inclusive) commuters drive to work alone?
c. What is the probability that more than 95 commuters drive to work alone?

6.52 According to figures released by the National Agricultural Statistics Service of the U.S. Department of Agriculture, the U.S. production of wheat over the past 20 years has been approximately uniformly distributed. Suppose the mean production over this period was 2.165 billion bushels. If the height of this distribution is .862 billion bushels, what are the values of a and b for this distribution?

6.53 The Federal Reserve System publishes data on family income based on its Survey of Consumer Finances. When the head of the household has a college degree, the mean before-tax family income is $70,400. Suppose that 60% of the before-tax family incomes when the head of the household has a college degree are between $61,200 and $79,600 and that these incomes are normally distributed. What is the standard deviation of before-tax family incomes when the head of the household has a college degree?

6.54 According to The Polk Company, a survey of households using the Internet in buying or leasing cars reported that 81% were seeking information about prices. In addition, 44% were seeking information about products offered. Suppose 75 randomly selected households are contacted who are using the Internet in buying or leasing cars.

a. What is the expected number of households who are seeking price information?
b. What is the expected number of households who are seeking information about products offered?
c. What is the probability that 67 or more households are seeking information about prices?
d. What is the probability that less than 23 households are seeking information about products offered?

6.55 Coastal businesses along the Gulf of Mexico from Texas to Florida worry about the threat of hurricanes during the season from June through October. Businesses become especially nervous when hurricanes enter the Gulf of Mexico. Suppose the arrival of hurricanes during this sea-
son is Poisson distributed, with an average of three hurricanes entering the Gulf of Mexico during the 5-month season. If a hurricane has just entered the Gulf of Mexico, what is the probability that at least 1 month will pass before the next hurricane enters the Gulf? What is the probability that another hurricane will enter the Gulf of Mexico in 2 weeks or less? What is the average amount of time between hurricanes entering the Gulf of Mexico?

6.56 With the growing emphasis on technology and the changing business environment, many workers are discovering that training such as reeducation, skill development, and personal growth are of great assistance in the job marketplace. A recent Gallup survey found that 80% of Generation Xers considered the availability of company-sponsored training as a factor to weigh in taking a job. If 50 Generation Xers are randomly sampled, what is the probability that fewer than 35 consider the availability of company-sponsored training as a factor to weigh in taking a job? What is the expected number? What is the probability that between 42 and 47 (inclusive) consider the availability of company-sponsored training as a factor to weigh in taking a job?

6.57 According to the Air Transport Association of America, the average operating cost of an MD-80 jet airliner is $2087 per hour. Suppose the operating costs of an MD-80 jet airliner are normally distributed with a standard deviation of $175 per hour. For what operating cost would only 20% of the operating costs be less? For what operating cost would 65% of the operating costs be more? What operating cost would be more than 85% of operating costs?

6.58 Supermarkets usually get very busy at about 5 p.m. on weekdays, because many workers stop by on the way home to shop. Suppose at that time arrivals at a supermarket’s express checkout station are Poisson distributed, with an average of .8 person/minute. If the clerk has just checked out the last person in line, what is the probability that at least 1 minute will elapse before the next customer arrives? Suppose the clerk wants to go to the manager’s office to ask a quick question and needs 2.5 minutes to do so. What is the probability that the clerk will get back before the next customer arrives?

6.59 According to Editor and Publisher Yearbook, the average daily circulation of The Wall Street Journal based on 1997 figures is 1,774,880. The standard deviation is 50,940. Assume the paper’s daily circulation is normally distributed. On what percentage of days would it surpass a circulation of 1,850,000? Suppose the paper cannot support the fixed expenses of a full-production setup if the circulation drops below 1,620,000. If the probability of this event occurring is low, the production manager might try to keep the full crew in place and not disrupt operations. How often will this event happen, based on the historical information?

6.60 Incoming phone calls generally are thought to be Poisson distributed. If an operator averages 2.2 phone calls every 30 seconds, what is the expected (average) amount of time between calls? What is the probability that a minute or more would elapse between incoming calls? Two minutes?

INTERPRETING THE OUTPUT

6.61 Shown here is a MINITAB output. Suppose the data represent the number of sales associates who are working in a department store in any given retail day. Describe the distribution including the mean and standard deviation. Interpret the shape of the distribution and the mean in light of the data being studied. What do the probability statements mean?

<table>
<thead>
<tr>
<th>Cumulative distribution function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous uniform on 11.0000 to 32.0000</td>
</tr>
<tr>
<td>( x )</td>
</tr>
<tr>
<td>28.0000</td>
</tr>
<tr>
<td>34.0000</td>
</tr>
<tr>
<td>16.0000</td>
</tr>
<tr>
<td>21.0000</td>
</tr>
</tbody>
</table>

6.62 A metal rod is being produced by a manufacturing company. Use the Excel output shown here to describe the weight of the rod. Interpret the probability values in terms of the manufacturing process.

| Normal Distribution: Mean = 227 mg, Standard Deviation = 2.3 mg |
|--------------------------|------------------|
| \( X \) Values | Probability \( \leq X \) |
| 220 | 0.0012 |
| 225 | 0.1923 |
| 227 | 0.5000 |
| 231 | 0.9590 |
| 238 | 1.0000 |

6.63 Suppose the MINITAB output shown here represents the analysis of the length of home-use cell phone calls in terms of minutes. Describe the distribution of cell phone call lengths and interpret the meaning of the probability statements.

<table>
<thead>
<tr>
<th>Cumulative distribution function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal with mean = 2.35000 and standard deviation = 0.11000</td>
</tr>
<tr>
<td>( x )</td>
</tr>
<tr>
<td>2.6000</td>
</tr>
<tr>
<td>2.4500</td>
</tr>
<tr>
<td>2.3000</td>
</tr>
<tr>
<td>2.0000</td>
</tr>
</tbody>
</table>

6.64 A restaurant averages 4.51 customers per 10 minutes during the summer in the late afternoon. Shown here are
Excel and MINITAB output for this restaurant. Discuss the type of distribution used to analyze the data and the meaning of the probabilities.

**Exponential Distribution Probabilities for**
**Lambda = 4.51**

<table>
<thead>
<tr>
<th>X Values</th>
<th>Probability &lt; X</th>
<th>Probability ≥ X</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.3630</td>
<td>0.6370</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5942</td>
<td>0.4058</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8951</td>
<td>0.1049</td>
</tr>
<tr>
<td>1</td>
<td>0.9890</td>
<td>0.0110</td>
</tr>
<tr>
<td>2.4</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Cumulative Distribution Function**

Exponential with mean = 0.221729

<table>
<thead>
<tr>
<th>x</th>
<th>P( X &lt;= x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1000</td>
<td>0.3630</td>
</tr>
<tr>
<td>0.2000</td>
<td>0.5942</td>
</tr>
<tr>
<td>0.5000</td>
<td>0.8951</td>
</tr>
<tr>
<td>1.0000</td>
<td>0.9890</td>
</tr>
<tr>
<td>2.4000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
CALCULATING THE STATISTICS

7.29 The mean of a population is 76 and the standard deviation is 14. The shape of the population is unknown. Determine the probability of each of the following occurring from this population.
   a. A random sample of size 35 yielding a sample mean of 79 or more
   b. A random sample of size 140 yielding a sample mean of between 74 and 77
   c. A random sample of size 219 yielding a sample mean of less than 76.5

7.30 Forty-six percent of a population possess a particular characteristic. Random samples are taken from this population. Determine the probability of each of the following occurrences.
   a. The sample size is 60 and the sample proportion is between .41 and .53
   b. The sample size is 458 and the sample proportion is less than .40
   c. The sample size is 1350 and the sample proportion is greater than .49

TESTING YOUR UNDERSTANDING

7.31 Suppose the age distribution in a city is as follows.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 18</td>
<td>22%</td>
</tr>
<tr>
<td>18–25</td>
<td>18%</td>
</tr>
<tr>
<td>26–50</td>
<td>36%</td>
</tr>
<tr>
<td>51–65</td>
<td>10%</td>
</tr>
<tr>
<td>Over 65</td>
<td>14%</td>
</tr>
</tbody>
</table>

A researcher is conducting proportionate stratified random sampling with a sample size of 250. Approximately how many people should he sample from each stratum?
7.32 Candidate Jones believes she will receive .55 of the total votes cast in her county. However, in an attempt to validate this figure, she has her pollster contact a random sample of 600 registered voters in the county. The poll results show that 298 of the voters say they are committed to voting for her. If she actually has .55 of the total vote, what is the probability of getting a sample proportion this small or smaller? Do you think she actually has 55% of the vote? Why or why not?

7.33 Determine a possible frame for conducting random sampling in each of the following studies.
   a. The average amount of overtime per week for production workers in a plastics company in Pennsylvania
   b. The average number of employees in all Alpha/Beta supermarkets in California
   c. A survey of commercial lobster catchers in Maine

7.34 A particular automobile costs an average of $17,755 in the Pacific Northwest. The standard deviation of prices is $650. Suppose a random sample of 30 dealerships in Washington and Oregon is taken and their managers are asked what they charge for this automobile. What is the probability of getting a sample average cost of less than $17,500? Assume that only 120 dealerships in the entire Pacific Northwest sell this automobile.

7.35 A company has 1250 employees, and you want to take a simple random sample of \( n = 60 \) employees. Explain how you would go about selecting this sample by using the table of random numbers. Are there numbers that you cannot use? Explain.

7.36 Suppose the average client charge per hour for out-of-court work by lawyers in the state of Iowa is $125. Suppose further that a random telephone sample of 32 lawyers in Iowa is taken and that the sample average charge per hour for out-of-court work is $110. If the population variance is $525, what is the probability of getting a sample mean this large or larger? What is the probability of getting a sample mean larger than $135 per hour? What is the probability of getting a sample mean of between $120 and $130 per hour?

7.37 A survey of 2645 consumers by DDB Needham Worldwide of Chicago for public relations agency Porter/Novelli showed that how a company handles a crisis when at fault is one of the top influences in consumer buying decisions, with 73% claiming it is an influence. Quality of product was the number-one influence, with 96% of consumers stating that quality has an influence on their buying decisions. How a company handles complaints was number two, with 85% of consumers reporting it as an influence in their buying decisions. Suppose a random sample of 1100 consumers is taken and each is asked which of these three factors influence their buying decisions.
   a. What is the probability that more than 810 consumers claim that how a company handles a crisis when at fault is an influence in their buying decisions?
   b. What is the probability that fewer than 1030 consumers claim that quality of product is an influence in their buying decisions?
   c. What is the probability that between 82% and 84% of consumers claim that how a company handles complaints is an influence in their buying decisions?

7.38 Suppose you are sending out questionnaires to a randomly selected sample of 100 managers. The frame for this study is the membership list of the American Managers Association. The questionnaire contains demographic questions about the company and its top manager. In addition, it asks questions about the manager’s leadership style. Research assistants are to score and enter the responses into the computer as soon as they are received. You are to conduct a statistical analysis of the data. Name and describe four nonsampling errors that could occur in this study.

7.39 A researcher is conducting a study of a Fortune 500 company that has factories, distribution centers, and retail outlets across the country. How can she use cluster or area sampling to take a random sample of employees of this firm?

7.40 A directory of personal computer retail outlets in the United States contains 12,080 alphabetized entries. Explain how systematic sampling could be used to select a sample of 300 outlets.

7.41 In an effort to cut costs and improve profits, many U.S. companies have been turning to outsourcing. In fact, according to Purchasing magazine, 54% of companies surveyed outsourced some part of their manufacturing process in the past two to three years. Suppose 565 of these companies are contacted.
   a. What is the probability that 339 or more companies have outsourced some part of their manufacturing process in the past two to three years?
   b. What is the probability that 288 or more companies have outsourced some part of their manufacturing process in the past two to three years?
   c. What is the probability that 50% or less of these companies have outsourced some part of their manufacturing process in the past two to three years?

7.42 The average cost of a one-bedroom apartment in a town is $550 per month. What is the probability of randomly selecting a sample of 50 one-bedroom apartments in this town and getting a sample mean of less than $530 if the population standard deviation is $100?

7.43 The Aluminum Association reports that the average American uses 56.8 pounds of aluminum in a year. A random sample of 51 households is monitored for 1 year to determine aluminum usage. If the population standard deviation of annual usage is 12.3 pounds, what is the probability that the sample mean will be each of the following?
   a. More than 60 pounds
   b. More than 58 pounds
c. Between 56 and 57 pounds
d. Less than 55 pounds
e. Less than 50 pounds

7.44 Use Table A.1 to select 20 three-digit random numbers. Did any of the numbers occur more than once? How can this happen? Make a stem and leaf plot of the numbers with the stem being the left digit. Do the numbers seem to be equally distributed, or are they bunched together?

7.45 Direct marketing companies are turning to the Internet for new opportunities. A recent study by Gruppo, Levey, & Co. showed that 73% of all direct marketers conduct transactions on the Internet. Suppose a random sample of 300 direct marketing companies is taken.

a. What is the probability that between 210 and 234 (inclusive) direct marketing companies are turning to the Internet for new opportunities?
b. What is the probability that 78% or more of direct marketing companies are turning to the Internet for new opportunities?
c. Suppose a random sample of 800 direct marketing companies is taken. Now what is the probability that 78% or more are turning to the Internet for new opportunities? How does this answer differ from the answer in part (b)? Why do the answers differ?

7.46 According to the U.S. Bureau of Labor Statistics, 20% of all people 16 years of age or older do volunteer work. Women volunteer slightly more than men, with 22% of women volunteering and 19% of men volunteering. What is the probability of randomly sampling 140 women 16 years of age or older and getting 35 or more who do volunteer work? What is the probability of getting 21 or fewer from this group? Suppose a sample of 300 men and women 16 years of age or older is selected randomly from the U.S. population. What is the probability that the sample proportion who do volunteer work is between 18% and 25%?

7.47 Suppose you work for a large firm that has 20,000 employees. The CEO calls you in and asks you to determine employee attitudes toward the company. She is willing to commit $100,000 to this project. What are the advantages of taking a sample versus conducting a census? What are the trade-offs?

7.48 In a particular area of the Northeast, an estimated 75% of the homes use heating oil as the principal heating fuel during the winter. A random telephone survey of 150 homes is taken in an attempt to determine whether this figure is correct. Suppose 120 of the 150 homes surveyed use heating oil as the principal heating fuel. What is the probability of getting a sample proportion this large or larger if the population estimate is true?

7.49 The U.S. Bureau of Labor Statistics released hourly wage figures for western countries in 1996 for workers in the manufacturing sector. The hourly wage was $28.34 in Switzerland, $20.84 in Japan, and $17.70 in the United States. Suppose 40 manufacturing workers are selected randomly from across Switzerland and asked what their hourly wage is. What is the probability that the sample average will be between $28 and $29? Suppose 35 manufacturing workers are selected randomly from across Japan. What is the probability that the sample average will exceed $22? Suppose 50 manufacturing workers are selected randomly from across the United States. What is the probability that the sample average will be less than $16.50? Assume that in all three countries, the standard deviation of hourly labor rates is $3.

7.50 Give a variable that could be used to stratify the population for each of the following studies. List at least four subcategories for each variable.

a. A political party wants to conduct a poll prior to an election for the office of U.S. senator in Minnesota.
b. A soft-drink company wants to take a sample of soft-drink purchases in an effort to estimate market share.
c. A retail outlet wants to interview customers over a 1-week period.
d. An eyeglasses manufacturer and retailer wants to determine the demand for prescription eyeglasses in its marketing region.

7.51 According to Runzheimer International, a typical business traveler spends an average of $281 per day in Chicago. This cost includes hotel, meals, car rental, and incidentals. A survey of 65 randomly selected business travelers who have been to Chicago on business recently is taken. For the population mean of $281 per day, what is the probability of getting a sample average of more than $273 per day if the population standard deviation is $47?
CALCULATING THE STATISTICS

8.47 Use the following data to construct 80%, 94%, and 98% confidence intervals to estimate \( \mu \). State the point estimate.

\[
44 \quad 37 \quad 49 \quad 30 \quad 56 \quad 48 \quad 53 \quad 42 \quad 51 \\
38 \quad 39 \quad 45 \quad 47 \quad 52 \quad 59 \quad 50 \quad 46 \quad 34 \\
39 \quad 46 \quad 27 \quad 35 \quad 52 \quad 51 \quad 46 \quad 45 \quad 58 \\
51 \quad 37 \quad 45 \quad 52 \quad 51 \quad 54 \quad 39 \quad 48
\]

8.48 Construct 90%, 95%, and 99% confidence intervals to estimate \( \mu \) from the following data. State the point estimate. Assume the data come from a normally distributed population.

\[
12.3 \quad 11.6 \quad 11.9 \quad 12.8 \quad 12.5 \\
11.4 \quad 12.0 \quad 11.7 \quad 11.8 \quad 12.3
\]

8.49 Use the following information to compute the confidence interval for the population proportion.

a. \( n = 715 \) and \( \hat{p} = 0.329 \) with 95% confidence
b. \( n = 284 \) and \( \hat{p} = 0.71 \) with 90% confidence
c. \( n = 1250 \) and \( \hat{p} = 0.48 \) with 95% confidence
d. \( n = 457 \) and \( \hat{p} = 0.27 \) with 98% confidence

8.50 Use the following data to construct 90% and 95% confidence intervals to estimate the population variance. Assume the data come from a normally distributed population.

\[
212 \quad 229 \quad 217 \quad 216 \quad 223 \quad 219 \quad 208 \quad 214 \quad 232 \quad 219
\]

8.51 Determine the sample size necessary under the following conditions.

a. To estimate \( \mu \) with \( \sigma = 44 \), \( E = 3 \), and 95% confidence
b. To estimate \( \mu \) with a range of values from 20 to 88 with \( W = 4 \) and 90% confidence
c. To estimate \( P \) with \( P \) unknown, \( E = 0.04 \), and 98% confidence

d. To estimate \( P \) with \( W = 0.06 \), 95% confidence, and \( P \) thought to be approximately 0.70

TESTING YOUR UNDERSTANDING

8.52 In planning both market opportunity and production levels, being able to estimate the size of a market can be important. Suppose a diaper manufacturer wants to know how many diapers a 1-month-old baby uses during a 24-hour period. To determine this, the manufacturer's analyst randomly selects 17 parents of 1-month-olds and asks them to keep track of diaper usage for 24 hours. The results are shown. Construct a 99% confidence interval to estimate the average daily diaper usage of a 1-month-old baby. Assume diaper usage is normally distributed.

\[
12 \quad 8 \quad 11 \quad 9 \quad 13 \quad 14 \quad 10 \\
10 \quad 7 \quad 13 \quad 11 \quad 8 \quad 11 \quad 15
\]

8.53 Suppose you want to estimate the proportion of cars that are sport utility vehicles (SUVs) being driven in Kansas City, Missouri, at rush hour by standing on the corner of I-70 and I-470 and counting SUVs. You believe the figure is no higher than 0.40. If you want the width of the confidence interval to be no greater than 0.06, how many cars should you randomly sample? Use a 90% level of confidence.

8.54 Use the data in Problem 8.52 to construct a 99% confidence interval to estimate the population variance for the number of diapers used during a 24-hour period for 1-month-olds. How could information about the population variance be used by a manufacturer or marketer in planning?

8.55 What is the average length of a company's policy book? Suppose policy books are sampled from 45 medium-size
companies. The average number of pages in the sample books is 213, with a sample standard deviation of 48. Use this information to construct a 98% confidence interval to estimate the mean number of pages for the population of medium-size company policy books.

8.56 A random sample of small-business managers was given a leadership style questionnaire. The results were scaled so that each manager received a score for initiative. Suppose the following data are a random sample of these scores.

<table>
<thead>
<tr>
<th>37</th>
<th>42</th>
<th>40</th>
<th>39</th>
<th>38</th>
<th>31</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>35</td>
<td>45</td>
<td>30</td>
<td>33</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>39</td>
<td>33</td>
<td>39</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>33</td>
<td>35</td>
<td>36</td>
<td>41</td>
<td>33</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>40</td>
<td>42</td>
<td>44</td>
<td>35</td>
<td>36</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>37</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Use these data to construct a 90% confidence interval to estimate the average score on initiative for all small-business managers.

8.57 A national beauty salon chain wants to estimate the number of times per year a woman has her hair done at a beauty salon if she uses one at least once a year. The chain's researcher estimates that, of those women who use a beauty salon at least once a year, the standard deviation of number of times of usage is approximately 6. The national chain wants to estimate the mean to be within one time of the actual mean value. How large a sample should the researcher take to obtain a 98% confidence level?

8.58 Is the environment a major issue with Americans? To answer that question, a researcher conducts a survey of 1255 randomly selected Americans. Suppose 714 of the sampled people replied that the environment is a major issue with them. Construct a 95% confidence interval to estimate the proportion of Americans who feel that the environment is a major issue with them. What is the point estimate of this proportion?

8.59 According to a survey by Topaz Enterprises, a travel auditing company, the average error by travel agents is $128. Suppose this figure was obtained from a random sample of 25 travel agents and the sample standard deviation is $21. What is the point estimate of the national average error for all travel agents? Compute a 98% confidence interval for the national average error based on these sample results. Assume the travel agent errors are normally distributed in the population. How wide is the interval? Interpret the interval.

8.60 A national survey on telemarketing was undertaken. One of the questions asked was: How long has your organization had a telemarketing operation? Suppose the following data represent some of the answers received to this question. Suppose further that only 300 telemarketing firms comprised the population when this survey was taken. Use the following data to compute a 98% confidence interval to estimate the average number of years a telemarketing organization has had a telemarketing operation.

<table>
<thead>
<tr>
<th>5</th>
<th>5</th>
<th>6</th>
<th>3</th>
<th>6</th>
<th>7</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td>5</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>14</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8.61 An entrepreneur wants to open an appliance service repair shop. She would like to know about what the average home repair bill is, including the charge for the service call for appliance repair in the area. She wants the estimate to be within $20 of the actual figure. She believes the range of such bills is between $30 and $600. How large a sample should the entrepreneur take if she wants to be 95% confident of the results?

8.62 A national survey of insurance offices was taken, resulting in a random sample of 245 companies. Of these 245 companies, 189 responded that they were going to purchase new software for their offices in the next year. Construct a 90% confidence interval to estimate the proportion of insurance offices that intend to purchase new software during the next year.

8.63 A national survey of companies included a question that asked whether the company had at least one bilingual telephone operator. The sample results of 90 companies follow (y denotes that the company does have at least one bilingual operator; n denotes that it does not).

<table>
<thead>
<tr>
<th>n</th>
<th>n</th>
<th>n</th>
<th>n</th>
<th>n</th>
<th>y</th>
<th>n</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
</tr>
</tbody>
</table>

Use this information to estimate with 95% confidence the proportion of the population that does have at least one bilingual operator.

8.64 A movie theater has had a poor accounting system. The manager has no idea how many large containers of popcorn are sold per movie showing. She knows that the amounts vary by day of the week and hour of the day. However, she wants to estimate the overall average per movie showing. To do so, she randomly selects 12 movie performances and counts the number of large containers of popcorn sold between 1/2 hour before the movie showing and 15 minutes after the movie showing.
sample average was 43.7 containers, with a variance of 228. Construct a 95% confidence interval to estimate the mean number of large containers of popcorn sold during a movie showing. Assume the number of large containers of popcorn sold per movie is normally distributed in the population. Use this information to construct a 98% confidence interval to estimate the population variance.

8.65 According to a survey by Runzheimer International, the average cost of a fast-food meal (quarter-pound cheeseburger, large fries, medium soft drink, excluding taxes) in Seattle is $4.82. Suppose this was based on a sample of 27 different establishments and the standard deviation was $0.37. Construct a 95% confidence interval for the population mean cost for all fast-food meals in Seattle. Assume the costs of a fast-food meal in Seattle are normally distributed. Using the interval as a guide, is it likely that the population mean is really $4.50? Why or why not?

8.66 A survey of 77 commercial airline flights of under 2 hours resulted in a sample average late time for a flight of 2.48 minutes. The sample standard deviation was 12 minutes. Construct a 95% confidence interval for the average time that a commercial flight of under 2 hours is late. What is the point estimate? What does the interval tell about whether the average flight is late?

8.67 A regional survey of 560 companies asked the vice president of operations how satisfied he or she was with the software support being received from the computer staff of the company. Suppose 33% of the 560 vice presidents said they were satisfied. Construct a 99% confidence interval for the proportion of the population of vice presidents who would have said they were satisfied with the software support if a census had been taken.

8.68 A research firm has been asked to determine the proportion of all restaurants in the state of Ohio that serve alcoholic beverages. The firm wants to be 98% confident of its results but has no idea of what the actual proportion is. The firm would like to report an error of no more than .05. How large a sample should it take?

8.69 A national magazine marketing firm attempts to win subscribers with a mail campaign that involves a contest using magazine stickers. Often when people subscribe to magazines in this manner they sign up for multiple magazine subscriptions. Suppose the marketing firm wants to estimate the average number of subscriptions per customer of those who purchase at least one subscription. To do so, the marketing firm’s researcher randomly selects 65 returned contest entries. Twenty-seven contain subscription requests. Of the 27, the average number of subscriptions is 2.10, with a standard deviation of .86. The researcher uses this information to compute a 98% confidence interval to estimate $\mu$ and assumes that $X$ is normally distributed. What does he find?

8.70 A national survey showed that Hillshire Farm Deli Select cold cuts were priced, on the average, at $5.20 per pound. Suppose a national survey of 23 retail outlets was taken and the price per pound of Hillshire Farm Deli Select cold cuts was ascertained. If the data below represent these prices, what is a 90% confidence interval for the population variance of these prices? Assume prices are normally distributed in the population.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.18</td>
<td>5.22</td>
<td>5.25</td>
<td>5.19</td>
<td>5.30</td>
<td></td>
</tr>
<tr>
<td>5.17</td>
<td>5.15</td>
<td>5.28</td>
<td>5.20</td>
<td>5.14</td>
<td></td>
</tr>
<tr>
<td>5.05</td>
<td>5.19</td>
<td>5.26</td>
<td>5.23</td>
<td>5.19</td>
<td></td>
</tr>
<tr>
<td>5.22</td>
<td>5.08</td>
<td>5.21</td>
<td>5.24</td>
<td>5.33</td>
<td></td>
</tr>
<tr>
<td>5.22</td>
<td>5.19</td>
<td>5.19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8.71 The price of a head of iceberg lettuce varies greatly with the season and the geographic location of a store. During February a researcher contacts a random sample of 39 grocery stores across the United States and asks the produce manager of each to state the current price charged for a head of iceberg lettuce. Using the researcher’s results that follow, construct a 99% confidence interval to estimate the mean price of a head of iceberg lettuce in February in the United States.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.59</td>
<td>$1.25</td>
<td>$1.65</td>
<td>$1.40</td>
<td>$0.89</td>
<td></td>
</tr>
<tr>
<td>1.19</td>
<td>1.50</td>
<td>1.49</td>
<td>1.30</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>1.29</td>
<td>1.60</td>
<td>0.99</td>
<td>1.29</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>1.50</td>
<td>1.49</td>
<td>1.29</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>1.10</td>
<td>0.89</td>
<td>1.10</td>
<td>1.39</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>1.50</td>
<td>1.55</td>
<td>1.20</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>1.00</td>
<td>1.30</td>
<td>1.25</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.55</td>
<td>1.29</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INTERPRETING THE OUTPUT**

8.72 A soft-drink company produces a cola in a 12 oz can. Even though their machines are set to fill the cans with 12 oz, variation due to calibration, operator error, and other things sometimes precludes the cans having the correct fill. To monitor the can fills, a quality team randomly selects some filled 12 oz cola cans and measures their fills in the lab. A confidence interval for the population mean is constructed from the data. Shown here is the MINITAB output from this effort. Discuss the output.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>99.0 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can Fills</td>
<td>58</td>
<td>11.9792</td>
<td>0.0536</td>
<td>0.0066</td>
<td>(.11.9623, 11.9962)</td>
</tr>
</tbody>
</table>

**Z CONFIDENCE INTERVALS**

The assumed sigma = 0.0500

Variable      N    Mean    StDev  SE Mean  99.0% CI
Can Fills     58  11.9792 0.0536  0.0066  (.11.9623, 11.9962)
A company has developed a new light bulb that seems to burn longer than most residential bulbs. To determine how long these bulbs burn, the company randomly selects a sample of these bulbs and burns them in the laboratory. The Excel output shown here is a portion of the analysis from this effort. Discuss the output.

8.74 Suppose a researcher wants to estimate the average age of a person who is a first-time home buyer. A random sample of first-time home buyers is taken and their ages are ascertained. The MINITAB output shown here is an analysis of that data. Study the output and explain its implication.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>98.0 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages</td>
<td>21</td>
<td>27.38</td>
<td>6.64</td>
<td>1.45</td>
<td>(23.72, 31.04)</td>
</tr>
</tbody>
</table>

8.75 What proportion of all American workers drive their cars to work? Suppose a poll of American workers is taken in an effort to answer that question, and the MINITAB output shown here is an analysis of the data from the poll. Explain the meaning of the output in light of the question.

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95.0 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>506</td>
<td>781</td>
<td>0.648</td>
<td>(0.613, 0.681)</td>
</tr>
</tbody>
</table>
SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

9.39 Use the information given and the eight-step process to test the hypotheses. Let $\alpha = .01$.

$H_0: \mu = 36 \quad H_1: \mu \neq 36$

$n = 63; \overline{X} = 38.4, S = 5.93$

9.40 Use the information given and the eight-step process to test the hypotheses. Let $\alpha = .05$. Assume the population is normally distributed.

$H_0: \mu = 7.82 \quad H_1: \mu < 7.82$

$n = 17; \overline{X} = 7.01, S = 1.69$

9.41 For each of the following problems, test the hypotheses. Incorporate the eight-step process.

a. $H_0: P = .28 \quad H_1: P > .28$

$n = 783, \overline{X} = 230, \alpha = .10$

b. $H_0: P = .61 \quad H_1: P \neq .61$

$n = 401, \hat{p} = .56, \alpha = .05$

9.42 Test the following hypotheses by using the information given and the eight-step process. Let alpha be .01. Assume the population is normally distributed.

$H_0: \sigma^2 = 15.4 \quad H_1: \sigma^2 > 15.4$

$n = 18, S^2 = 29.6$

9.43 Solve for the value of beta in each of the following problems.

a. $H_0: \mu = .130; H_1: \mu > .130; n = 75, \sigma = 12, \alpha = .01$. The alternative mean is actually 135.

b. $H_0: P = .44; H_1: P < .44; n = 1095, \alpha = .05$. The alternative proportion is actually .42.

TESTING YOUR UNDERSTANDING

9.44 According to a survey by ICR for Vienna Systems, a majority of American households have tried to cut long-distance phone bills. Of those who have tried to cut the bills, 32% have done so by switching long-distance companies. Suppose that lately there has been a frenzy of "slamming" (where the customer's long-distance provider is switched without the customer's knowledge or approval) and long-distance company solicitation and advertising. Because of this, ICR conducts another survey by randomly contacting 80 American households who have tried to cut long-distance phone bills. If 39% of the contacted households say they have tried to cut their long-distance phone bills by switching long-distance companies, is this enough evidence to state that a significantly higher proportion of American households are trying to cut long-distance phone bills by switching companies? Let $\alpha = .01$.

9.45 According to Zero Population Growth, the average urban U.S. resident uses 3.3 pounds of food per day. Is this figure accurate for rural U.S. residents? Suppose 64 rural U.S. residents are identified by a random procedure and their average consumption per day is 3.45 pounds of food. Assume a population variance of 1.31 pounds of food per day. Use a 5% level of significance to determine whether the Zero Population Growth figure for urban U.S. residents also is true for rural U.S. residents on the basis of the sample data.

9.46 Brokers generally agree that bonds are a better investment during times of low interest rates than during times of high interest rates. A survey of executives during a time of low interest rates showed that 57% of them had some retirement funds invested in bonds. Assume this percentage is constant for bond market investment by executives with retirement funds. Suppose interest rates have risen lately and the proportion of executives with retirement investment money in the bond market may have dropped. To test this idea, a researcher randomly samples 210 executives who have retirement funds. Of these, 93 now have retirement funds invested in bonds. For $\alpha = .10$, is this enough evidence to declare that the proportion of executives with retirement fund investments in the bond market is significantly lower than .57?

9.47 Highway engineers in Ohio are painting white stripes on a highway. The stripes are supposed to be approximately 10 feet long. However, because of the machine, the operator, and the motion of the vehicle carrying the equipment, there is considerable variation among the stripe lengths. Engineers claim that the variance of stripes is not more than 16 inches. Use the sample lengths given here from 12 measured stripes to test the variance claim. Assume stripe length is normally distributed. Let $\alpha = .05$.

<table>
<thead>
<tr>
<th>STRIPE LENGTHS IN FEET</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.3</td>
</tr>
<tr>
<td>9.2</td>
</tr>
<tr>
<td>9.3</td>
</tr>
</tbody>
</table>

9.48 A computer manufacturer estimates that its line of minicomputers has, on average, 8.4 days of downtime per year. To test this claim, a researcher contacts seven companies that own one of these computers and is allowed to
access company computer records. It is determined that, for the sample, the average number of downtime days is 5.6, with a sample standard deviation of 1.3 days. Assuming that number of downtime days is normal, test to determine whether these minicomputers actually average 8.4 days of downtime in the entire population. Let $\alpha = .01$.

9.49 A life insurance salesperson claims the average worker in the city of Cincinnati has no more than $25,000 of personal life insurance. To test this claim, you randomly sample 100 workers in Cincinnati. You find that this sample of workers has an average of $26,650 of personal life insurance and that the standard deviation is $12,000.

a. Determine whether there is enough evidence to reject the null hypothesis posed by the salesperson. Assume the probability of committing a Type I error is .05.

b. If the actual average for this population is $30,000, what is the probability of committing a Type II error?

9.50 A financial analyst has been watching a particular stock for several months. It appears that the price of this stock has remained fairly stable during this time. In fact, the financial analyst claims that the variance of the price of this stock has not exceeded $4 for the entire period. Recently, the market has heated up, and it appears that the price of this stock is more volatile. To determine whether this is so, a sample of closing prices of this stock for 8 days is taken randomly. The sample mean price is $36.25, with a sample standard deviation of $7.80. Using a level of significance of .10, test to determine whether the financial analyst's previous variance figure is now too low. Assume stock prices are normally distributed.

9.51 A study of MBA graduates by Universum for The American Graduate Survey 1999 revealed that MBA graduates have several expectations from prospective employers beyond their base pay. In particular, according to the study 46% expect a performance-related bonus, 46% expect stock options, 42% expect a signing bonus, 28% expect profit sharing, 27% expect extra vacation/personal days, 25% expect tuition reimbursement, 24% expect health benefits, and 19% expect guaranteed annual bonuses. Suppose a study is conducted in an ensuing year to see if these expectations have changed. If 125 MBA graduates are randomly selected and if 66 expect stock options, is this enough evidence to declare that a significantly higher proportion of MBAs expect stock options? Let $\alpha = .05$. If the proportion really is .50, what is the probability of committing a Type II error?

9.52 Suppose the number of beds filled per day in a medium-size hospital is normally distributed. A hospital administrator has been quoted as having told the board of directors that, on the average, at least 185 beds are filled on any given day. One of the board members believes this figure is inflated, and she manages to secure a random sample of figures for 16 days. The data are shown here. Use $\alpha = .05$ and the sample data to test whether the hospital administrator's statement is false. Assume the number of filled beds per day is normally distributed in the population.

<table>
<thead>
<tr>
<th>NUMBER OF BEDS OCCUPIED PER DAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>173</td>
</tr>
<tr>
<td>189</td>
</tr>
<tr>
<td>177</td>
</tr>
<tr>
<td>199</td>
</tr>
</tbody>
</table>

9.53 According to the International Data Corporation, Compaq Computers holds a 16% share of the personal computer market in the United States and a 12.7% share of the worldwide market. Suppose a market researcher believes that Compaq holds a higher share of the market in the southwestern region of the United States. To verify this theory, he randomly selects 428 people who have purchased a personal computer in the last month in the southwestern region of the United States. Eighty-four of these purchases were Compaq Computers. Using a 1% level of significance, test the market researcher's theory. What is the probability of making a Type I error? If the market share is really .21 in the southwestern region of the United States, what is the probability of making a Type II error?

9.54 A national publication reported that a college student living away from home spends, on average, no more than $15 per month on laundry. You believe this figure is too low and want to disprove this claim. To conduct the test, you randomly select 35 college students and ask them to keep track of the amount of money they spend during a given month for laundry. The sample produces an average expenditure on laundry of $19.34, with a standard deviation of $4.52. Use these sample data to conduct the hypothesis test. Assume you are willing to take a 10% risk of making a Type I error.

9.55 A local company installs natural-gas grills. As part of the installation, a ditch is dug to lay a small natural-gas line from the grill to the main line. On the average, the depth of these lines seems to run about 1 foot. The company claims that the depth does not vary by more than 16 inches (the variance). To test this claim, a researcher randomly took 22 depth measurements at different locations. The sample average depth was 13.4 inches with a standard deviation of 6 inches. Is this enough evidence to reject the company's claim about the variance? Assume line depths are normally distributed. Let $\alpha = .05$.

9.56 A study of pollutants showed that certain industrial emissions should not exceed 2.5 parts per million. You believe
9.57 The average cost per square foot for office rental space in the central business district of Philadelphia is $23.58, according to Cushman & Wakefield, Inc. A large real estate company wants to confirm this figure. The firm conducts a telephone survey of 95 offices in the central business district of Philadelphia and asks the office managers how much they pay in rent per square foot. Suppose the sample average is $22.83 per square foot, with a standard deviation of $5.11.

a. Conduct a hypothesis test using $\alpha = .05$ to determine whether the cost per square foot reported by Cushman & Wakefield, Inc. should be rejected.

b. If the decision in part (a) is to fail to reject and if the actual average cost per square foot is $22.30, what is the probability of committing a Type II error?

9.58 The American Water Works Association reports that, on average, men use between 10 and 15 gallons of water daily to shave when they leave the water running. Suppose the following data are the numbers of gallons of water used in a day to shave by 12 randomly selected men and the data come from a normal distribution of data. Use these data and a 5% level of significance to test to determine whether the population variance for such water usage is 2.5 gallons.

9.60 One survey conducted by RHI Management Resources determined that the Lexus is the favorite luxury car for 25% of CFOs. Suppose a financial management association conducts its own survey of CFOs in an effort to determine if this figure is correct. They use an alpha of .05. Following is the MINITAB output with the results of the survey. Discuss the findings, including the hypotheses, one- or two-tailed test, sample statistics, and the conclusion. Explain from the data why you reached the conclusion you did.

### Test and Confidence Interval for One Proportion

<table>
<thead>
<tr>
<th>Test of $p = 0.25$ vs $p \neq 0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exact</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample X</th>
<th>N</th>
<th>Sample $p$</th>
<th>99.0% CI</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79</td>
<td>0.205729</td>
<td>(0.155244, 0.263701)</td>
<td>0.045</td>
</tr>
</tbody>
</table>

9.61 In a recent year, published statistics by the National Cattlemen's Beef Association claimed that the average retail beef price for USDA All Fresh beef was $2.51. Suppose a survey of retailers is conducted this year to determine whether the price of USDA All Fresh beef has increased. The Excel output of the results of the survey are shown here. Analyze the output and explain what it means in this study.

### T-TEST

<table>
<thead>
<tr>
<th>Variable 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>t Stat</td>
</tr>
<tr>
<td>$P(T&lt;=t)$ one-tail</td>
</tr>
<tr>
<td>$t$ Critical one-tail</td>
</tr>
<tr>
<td>$P(T&lt;=t)$ two-tail</td>
</tr>
<tr>
<td>$t$ Critical two-tail</td>
</tr>
</tbody>
</table>

9.62 The American Express Retail Index states that the average U.S. household will spend $2,747 on home-improvement projects this year. Suppose a large national home-improvement company wants to test that figure in the West, theorizing that the average might be lower in the West. The research firm hired to conduct the study arrive at the results shown here. Analyze the data and explain the results.

### Z-TEST

<table>
<thead>
<tr>
<th>Test of $\mu = 2,747$ vs $\mu &lt; 2,747$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The assumed sigma = 1557</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>T</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home Imp</td>
<td>67</td>
<td>2327</td>
<td>1557</td>
<td>190</td>
<td>-2.21</td>
<td>0.014</td>
</tr>
</tbody>
</table>

**INTERPRETING THE OUTPUT**

9.59 According to the U.S. Bureau of the Census, the average American generates 4.4 pounds of garbage per day. Suppose we believe that because of recycling and a greater emphasis on the environment, the figure is now lower. To test this notion, we take a random sample of Americans and have them keep a log of their garbage for a day. We record and analyze the results by using a statistical computer package. The output is shown below. Describe the sample. What statistical decisions can be made on the basis of this analysis? Let alpha be .05.

**T-Test of the Mean**

Test of $\mu = 4.400$ vs $\mu < 4.400$

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>T</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>garbage</td>
<td>22</td>
<td>4.030</td>
<td>0.920</td>
<td>0.196</td>
<td>-1.89</td>
<td>0.037</td>
</tr>
</tbody>
</table>
CALCULATING THE STATISTICS

10.45 Test the following hypotheses with the data given. Let $\alpha = .10$.

$$H_0: \mu_1 - \mu_2 = 0 \quad H_A: \mu_1 - \mu_2 \neq 0$$

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 138.4$</td>
<td>$\bar{X}_1 = 142.5$</td>
</tr>
<tr>
<td>$S_1 = 6.71$</td>
<td>$S_2 = 8.92$</td>
</tr>
<tr>
<td>$n_1 = 48$</td>
<td>$n_2 = 39$</td>
</tr>
</tbody>
</table>

10.46 Use the data below to construct a 98% confidence interval to estimate the difference between $\mu_1$ and $\mu_2$.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 34.9$</td>
<td>$\bar{X}_2 = 27.6$</td>
</tr>
<tr>
<td>$S_1^2 = 2.97$</td>
<td>$S_2^2 = 3.50$</td>
</tr>
<tr>
<td>$n_1 = 34$</td>
<td>$n_2 = 31$</td>
</tr>
</tbody>
</table>

10.47 The following data come from independent samples drawn from normally distributed populations. Use these data to test the following hypotheses. Let the Type I error rate be .05.

$$H_0: \mu_1 - \mu_2 = 0 \quad H_A: \mu_1 - \mu_2 > 0$$

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 2.06$</td>
<td>$\bar{X}_2 = 1.93$</td>
</tr>
<tr>
<td>$S_1^2 = .176$</td>
<td>$S_2^2 = .143$</td>
</tr>
<tr>
<td>$n_1 = 12$</td>
<td>$n_2 = 15$</td>
</tr>
</tbody>
</table>

10.48 Construct a 95% confidence interval to estimate $\mu_1 - \mu_2$ by using the following data. Assume the populations are normally distributed.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 74.6$</td>
<td>$\bar{X}_2 = 70.9$</td>
</tr>
<tr>
<td>$S_1^2 = 10.5$</td>
<td>$S_2^2 = 11.4$</td>
</tr>
<tr>
<td>$n_1 = 18$</td>
<td>$n_2 = 19$</td>
</tr>
</tbody>
</table>

10.49 The following data have been gathered from two related samples. The differences are assumed to be normally distributed in the population. Use these data and alpha of .01 to test the following hypotheses.

$$H_0: D = 0 \quad H_A: D < 0$$

$n = 21, \bar{d} = -1.16, S_d = 1.01$

10.51 Test the following hypotheses by using the given data and alpha equal to .05.

$$H_0: P_1 - P_2 = 0 \quad H_A: P_1 - P_2 \neq 0$$

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1 = 783$</td>
<td>$n_2 = 896$</td>
</tr>
<tr>
<td>$X_1 = 345$</td>
<td>$X_2 = 421$</td>
</tr>
</tbody>
</table>

10.52 Use the following data to construct a 99% confidence interval to estimate $P_1 - P_2$.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1 = 409$</td>
<td>$n_2 = 378$</td>
</tr>
<tr>
<td>$\hat{p}_1 = .71$</td>
<td>$\hat{p}_2 = .67$</td>
</tr>
</tbody>
</table>

10.53 Test the following hypotheses by using the given data. Let alpha = .05.

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_A: \sigma_1^2 \neq \sigma_2^2$$

$n_1 = 8, n_2 = 10, S_1^2 = 46, S_2^2 = 37$

TESTING YOUR UNDERSTANDING

10.54 Suppose a large insurance company wants to estimate the difference between the average amount of term life insurance purchased per family and the average amount of whole life insurance purchased per family. To obtain an estimate, one of the company’s actuaries randomly selects 27 families who have term life insurance only and 29 families who have whole life policies only. Each sample is taken from families in which the leading provider is younger than 45 years of age. Use the data obtained to construct a 95% confidence interval to estimate the difference in means for these two groups. Assume the amount of insurance is normally distributed.

<table>
<thead>
<tr>
<th>TERM</th>
<th>WHOLE LIFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_t = 75,000$</td>
<td>$\bar{X}_w = 45,000$</td>
</tr>
<tr>
<td>$S_t = 22,000$</td>
<td>$S_w = 15,500$</td>
</tr>
<tr>
<td>$n_t = 27$</td>
<td>$n_w = 29$</td>
</tr>
</tbody>
</table>
A study is conducted to estimate the average difference in bus ridership for a large city during the morning and afternoon rush hours. The transit authority's researcher randomly selects nine buses because of the variety of routes they represent. On a given day the number of riders on each bus is counted at 7:45 A.M. and at 4:45 P.M., with the following results.

<table>
<thead>
<tr>
<th>BUS</th>
<th>MORNING</th>
<th>AFTERNOON</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>44</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>46</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>51</td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>49</td>
</tr>
</tbody>
</table>

Use the data to compute a 90% confidence interval to estimate the population average difference. Assume ridership is normally distributed.

According to CardWeb Inc.'s CardData, Visa's share of the U.S. credit card market went from 44.7% in 1990 to 48.7% in 1997. How were these figures obtained? Possibly they were determined by examining the total dollar amounts of transactions for the population of credit card usage and determining Visa's portion. If so, certainly we could declare that Visa's market share increased. Suppose, however, that the 1990 figure was obtained by randomly selecting 1300 credit card transactions and tallying the proportion done with a Visa card. Suppose also that the 1997 figure was obtained by randomly selecting 1450 credit card transactions and doing the same thing. At a 5% level of significance, is there enough evidence in the data taken from the samples to declare that Visa's market share increased significantly between 1990 and 1997?

A study was conducted to compare the salaries of accounting clerks and data entry operators. One of the hypotheses to be tested is that the variability of salaries among accounting clerks is the same as the variability of salaries of data entry operators. To test this, a random sample of 16 accounting clerks was taken, resulting in a sample mean salary of $26,400 and a sample standard deviation of $1200. A random sample of 14 data entry operators was taken as well, resulting in a sample mean of $25,800 and a sample standard deviation of $1050. Use these data and $\alpha = .05$ to test to determine whether the population variance of salaries is the same for accounting clerks as it is for data entry operators.

Is there more variation in the output of one shift in a manufacturing plant than in another shift? In an effort to study this question, plant managers gathered productivity reports from the 8 A.M. to 4 P.M. shift for 8 days. The reports indicated that the following numbers of units were produced on each day for this shift.

5528 4779 5112 5380 4918 4763 5055 5106

Productivity information was also gathered from 7 days for the 4 P.M. to midnight shift, resulting in the following data.

4325 4016 4872 4559 3982 4754 4116

Use these data and $\alpha = .01$ to test to determine whether the variances of productivity for the two shifts are the same. Assume productivity is normally distributed in the population.

A study was conducted to develop a scale to measure stress in the workplace. Respondents were asked to rate 26 distinct work events. Each event was to be compared with the stress of the first week on the job, which was awarded an arbitrary score of 500. Sixty professional men and 41 professional women participated in the study. One of the stress events was "lack of support from the boss." The men's sample average rating of this event was 631 and the women's sample average rating was 548. Suppose the sample standard deviations for men and for women both were about 100. Construct a 95% confidence interval to estimate the difference in the population mean scores on this event for men and women.

A national grocery store chain wants to estimate the difference in the average weight of turkeys sold in Detroit and the average weight of turkeys sold in Charlotte. According to the chain's researcher, a random sample of 20 turkeys sold at the chain's stores in Detroit yielded a sample mean of 17.53 pounds, with a standard deviation of 3.2 pounds. Her random sample of 24 turkeys sold at the chain's stores in Charlotte yielded a sample mean of 14.89 pounds, with a standard deviation of 2.7 pounds. Use a 1% level of significance to determine whether there is a difference in the mean weight of turkeys sold in these two cities. Assume the population variances are approximately the same and that the weights of turkeys sold in the stores are normally distributed.

A tree nursery has been experimenting with fertilizer to increase the growth of seedlings. A sample of 35 two-year-old pine trees are grown for three more years with a cake of fertilizer buried in the soil near the trees' roots. A second sample of 35 two-year-old pine trees are grown for three more years under identical conditions (soil, temperature, water) as the first group, but not fertilized. Tree growth is measured over the 3-year period with the following results.

<table>
<thead>
<tr>
<th>TREES WITH FERTILIZER</th>
<th>TREES WITHOUT FERTILIZER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1 = 35$</td>
<td>$n_2 = 35$</td>
</tr>
<tr>
<td>$\bar{X}_1 = 38.4$ inches</td>
<td>$\bar{X}_2 = 23.1$ inches</td>
</tr>
<tr>
<td>$S_1 = 9.8$ inches</td>
<td>$S_2 = 7.4$ inches</td>
</tr>
</tbody>
</table>

Do the data support the theory that the population of trees with the fertilizer grew significantly larger during the period in which they were fertilized than the nonfertilized trees? Use $\alpha = .01$. 

10.62 One of the most important aspects of a store’s image is the perceived quality of its merchandise. Other factors include merchandise pricing, assortment of products, convenience of location, and service. Suppose image perceptions of shoppers of specialty stores and shoppers of discount stores are being compared. A random sample of shoppers is taken at each type of store, and the shoppers are asked whether the quality of merchandise is a determining factor in their perception of the store’s image. Some 75% of the 350 shoppers at the specialty stores say yes, but only 52% of the 500 shoppers at the discount store say yes. Construct a 90% confidence interval for the difference in population proportions.

10.63 Use the data from Problem 10.54 to determine whether the variances of term and whole life insurance amounts are the same. Let $\alpha = .05$.

10.64 What is the average difference between the price of name-brand soup and the price of store-brand soup? To obtain an estimate, an analyst randomly samples eight stores. Each store sells its own brand and a national name brand. The prices of a can of name-brand tomato soup and a can of the store-brand tomato soup follow.

<table>
<thead>
<tr>
<th>STORE</th>
<th>NAME BRAND</th>
<th>STORE BRAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54¢</td>
<td>49¢</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>59</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>61</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>51</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>49</td>
</tr>
</tbody>
</table>

Construct a 90% confidence interval to estimate the average difference. Assume the prices of tomato soup are normally distributed in each population.

10.65 As the prices of heating oil and natural gas increase, consumers become more careful about heating their homes. Researchers want to know how warm homeowners keep their houses in January and how the results from Wisconsin and Tennessee compare. The researchers randomly call 23 Wisconsin households between 7 P.M. and 9 P.M. on January 15 and ask the respondent how warm the house is according to the thermostat. The researchers then call 19 households in Tennessee the same night and ask the same question. The results follow.

<table>
<thead>
<tr>
<th>WISCONSIN</th>
<th>TENNESSEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>71</td>
</tr>
<tr>
<td>70</td>
<td>61</td>
</tr>
<tr>
<td>75</td>
<td>68</td>
</tr>
<tr>
<td>74</td>
<td>68</td>
</tr>
<tr>
<td>69</td>
<td>72</td>
</tr>
<tr>
<td>70</td>
<td>73</td>
</tr>
</tbody>
</table>

For $\alpha = .01$, is the average temperature of a house in Tennessee significantly higher than that of a house in Wisconsin on the evening of January 15? Assume the population variances are equal and the house temperatures are normally distributed in each population.

10.66 In manufacturing, does worker productivity drop on Friday? In an effort to determine whether it does, a company’s personnel analyst randomly selects from a manufacturing plant five workers who make the same part. He measures their output on Wednesday and again on Friday and obtains the following results.

<table>
<thead>
<tr>
<th>WORKER</th>
<th>WEDNESDAY OUTPUT</th>
<th>FRIDAY OUTPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>68</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>74</td>
<td>58</td>
</tr>
</tbody>
</table>

The analyst uses $\alpha = .05$ and assumes the difference in productivity is normally distributed. Is there enough evidence to show that productivity drops on Friday?

10.67 A manufacturer has two machines that drill holes in pieces of sheet metal used in engine construction. The workers who attach the sheet metal to the engine become inspectors in that they reject sheets that have been so poorly drilled that they cannot be attached. The production manager is interested in knowing whether one machine produces more defective drillings than the other machine. As an experiment, employees mark the sheets so that the manager can determine which machine was used to drill the holes. A random sample of 191 sheets of metal drilled by machine 1 is taken, and 38 of the sheets are defective. A random sample of 202 sheets of metal drilled by machine 2 is taken, and 21 of the sheets are defective. Use $\alpha = .05$ to determine whether there is a significant difference in the proportion of sheets drilled with defective holes between machine 1 and machine 2.

10.68 Is there a difference in the proportion of construction workers who are under 35 years of age and the proportion of telephone repair people who are under 35 years of age? Suppose a study is conducted in Calgary, Alberta, using random samples of 338 construction workers and 281 telephone repair people. The sample of construction workers includes 297 people under 35 years of age and the sample of telephone repair people includes 192 people under that age. Use these data to construct a 90% confidence interval to estimate the difference in proportions of people under 35 years of age among construction workers and telephone repair people.

10.69 Executives often spend so many hours in meetings that they have relatively little time to manage their individual areas of operation. What is the difference in mean time spent in meetings by executives of the aerospace industry and executives of the automobile industry? Suppose random samples of 33 aerospace executives and
35 automobile executives are monitored for a week to determine how much time they spend in meetings. The results follow.

\[
\begin{array}{cccc}
\text{AEROSPACE} & \text{AUTOMOBILE} \\
\hline
n_1 = 33 & n_2 = 35 \\
\bar{X}_1 = 12.4 \text{ hours} & \bar{X}_2 = 4.6 \text{ hours} \\
S_1 = 2.9 \text{ hours} & S_2 = 1.8 \text{ hours} \\
\end{array}
\]

Use the data to estimate the difference in the mean time per week executives in these two industries spend in meetings. Use a 99% level of confidence.

10.70 Various types of retail outlets sell toys during the holiday season. Among them are specialty toy stores, the large discount toy stores, and other retailers that carry toys as only one part of their stock of goods. Is there any difference in the dollar amount of a customer purchase between a large discount toy store and a specialty toy store if they carry relatively comparable types of toys? Suppose in December a random sample of 60 sales slips is selected from a large discount toy outlet and a random sample of 40 sales slips is selected from a specialty toy store. The data gathered from these samples follow.

\[
\begin{array}{cccc}
\text{LARGE DISCOUNT} & \text{SPECIALTY} \\
\text{TOY STORE} & \text{TOY STORE} \\
\hline
\bar{X}_a = 47.20 & \bar{X}_s = 27.40 \\
S_a = 12.45 & S_s = 9.82 \\
\end{array}
\]

Use \( \alpha = .01 \) and the data to determine whether there is a significant difference in the average size of purchases at these stores.

10.71 One of the new thrusts of quality control management is to examine the process by which a product is produced. This approach also applies to paperwork. In industries where large long-term projects are undertaken, days and even weeks may elapse as a change order makes its way through a maze of approvals before receiving final approval. This process can result in long delays and stretch schedules to the breaking point. Suppose a quality control consulting group claims that it can significantly reduce the number of days required for such paperwork to receive approval. In an attempt to "prove" its case, the group selects five jobs for which it revises the paperwork system. The following data show the number of days required for a change order to be approved before the group intervened and the number of days required for a change order to be approved after the group instituted a new paperwork system.

\[
\begin{array}{cccc}
\text{BEFORE} & \text{AFTER} \\
\hline
12 & 8 \\
7 & 3 \\
10 & 8 \\
16 & 9 \\
8 & 5 \\
\end{array}
\]

Use \( \alpha = .01 \) to determine whether there was a significant drop in the number of days required to process paperwork to approve change orders. Assume that in each case the number of days of paperwork is normally distributed.

10.72 There are two large newspapers in your city. You are interested in knowing whether there is a significant difference in the average number of pages in each newspaper dedicated solely to advertising. You randomly select 10 editions of newspaper A and six editions of newspaper B (excluding weekend editions). The data follow. Use \( \alpha = .01 \) to test whether there is a significant difference in averages. Assume the number of pages of advertising per edition is normally distributed and the population variances are approximately equal.

\[
\begin{array}{cccc}
A & B \\
17 & 17 \\
21 & 15 \\
11 & 19 \\
19 & 22 \\
26 & 16 \\
\end{array}
\]

INTERPRETING THE OUTPUT

10.73 A study by Colliers International presented the highest and the lowest global rental rates per year per square foot of office space. Among the cities with the lowest rates were Perth, Australia; Edmonton, Alberta, Canada; and Calgary, Alberta, Canada with rates of $8.81, $9.55, and $9.69, respectively. At the high end were Hong Kong; Bombay, India; and Tokyo, Japan with rates over $100. Suppose a researcher conducted her own survey of businesses renting office space to determine whether one city is significantly more expensive than another. The data are tallied and analyzed by using MINITAB. The results follow. Discuss the output. What cities were studied? How large were the samples? What were the sample statistics? What was the value of alpha? What were the hypotheses and what was the conclusion?

<table>
<thead>
<tr>
<th>Two Sample T-Test and Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>Hong Kong</td>
</tr>
<tr>
<td>Bombay</td>
</tr>
</tbody>
</table>

98% C.I. For \( \mu_{\text{Hong Kong}} - \mu_{\text{Bombay}} \): \((-1.5, 18.4)\).

\[ T = 2.06 \text{ (vs >)} \]
\[ P = 0.023 \text{ DF } = 40 \]

Both use Pooled StDev = 13.2

10.74 Why do employees "blow the whistle" on other employees for unethical or illegal behavior? One study conducted by the AICPA reported the likelihood that employees would blow the whistle on another employee
for such things as unsafe working conditions, unsafe products, and poorly managed operations. On a scale from 1 to 7, with 1 denoting highly improbable and 7 denoting highly probable, unnecessary purchases received a 5.72 in the study. Suppose this study was administered at a company and then all employees were subjected to a one-month series of seminars on reasons to blow the whistle on fellow employees. One month later the study was administered again to the same employees at the company in an effort to determine whether the treatment had any effect. The following Excel output shows the results of the study. What were the sample sizes? What might the hypotheses have been? If \( \alpha = 0.05 \), what conclusions could be made? Which of the statistical tests presented in this chapter is likely to have been used?

**t-TEST: PAIRED TWO SAMPLE FOR MEANS**

<table>
<thead>
<tr>
<th>Variable 1</th>
<th>Variable 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.863</td>
</tr>
<tr>
<td>Variance</td>
<td>0.716</td>
</tr>
<tr>
<td>Observations</td>
<td>12</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.39</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>11</td>
</tr>
<tr>
<td>t Stat</td>
<td>-3.38</td>
</tr>
<tr>
<td>P(T&lt;)=t one-tail</td>
<td>0.0031</td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1.80</td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.0062</td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.20</td>
</tr>
</tbody>
</table>

10.76 A manufacturing company produces plastic pipes that are specified to be 10 inches long and 1/8 inch thick with an opening of 3/4 inch. These pipes are molded on two different machines. To maintain consistency, the company periodically randomly selects pipes for testing. In one specific test, pipes were randomly sampled from each machine and the lengths were measured. A statistical test was computed using Excel in an effort to determine whether there was a significant difference in the variances of the lengths of the pipes produced by the two machines. The results are shown here. Discuss the outcome of this test along with some of the other information given in the output.

**F-TEST TWO-SAMPLE FOR VARIANCES**

<table>
<thead>
<tr>
<th>Machine A</th>
<th>Machine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>10.031</td>
</tr>
<tr>
<td>Variance</td>
<td>0.04031</td>
</tr>
<tr>
<td>Observations</td>
<td>26</td>
</tr>
<tr>
<td>df</td>
<td>25</td>
</tr>
<tr>
<td>F</td>
<td>2.06</td>
</tr>
<tr>
<td>P(F&lt;=f) one-tail</td>
<td>0.0348</td>
</tr>
<tr>
<td>F Critical one-tail</td>
<td>1.92</td>
</tr>
</tbody>
</table>

10.75 A large manufacturing company produces computer printers that are distributed and sold all over the United States. Due to lack of industry information, the company has a difficult time ascertaining its market share in different parts of the country. They hire a market research firm to estimate their market share in a northern city and a southern city. They would also like to know if there is a difference in their market shares in these two cities; if so, they want to estimate how much. The market research firm randomly selects printer cus-

tomers from different locales across both cities and determines what brand of computer printer they purchased. The following MINITAB output shows the results from this study. Discuss the results including sample sizes, estimation of the difference in proportions, and any significant differences determined. What were the hypotheses tested?

**TEST AND CONFIDENCE INTERVAL FOR TWO PROPORTIONS**

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern City</td>
<td>147</td>
<td>473</td>
<td>0.310782</td>
</tr>
<tr>
<td>Southern City</td>
<td>104</td>
<td>385</td>
<td>0.270130</td>
</tr>
</tbody>
</table>

Estimate for \( p(1) - p(2) \): 0.0406524
99\% CI for \( p(1) - p(2) \): (-0.0393623, 0.120667)
Test for \( p(1) - p(2) = 0 \) (vs not = 0):
\[ Z = 1.31 \quad P\text{-Value} = 0.191 \]

10.10 The variable Service in the hospital database differentiates general medical hospitals (coded 1) and psychiatric hospitals (coded 2). Now test to determine whether there is a difference between these two types of hospitals on the variables Beds and Total Expenses.

4. Use the financial database to test to determine whether there is a significant difference in the proportion of companies whose earnings per share are more than $2.00 and the proportion of companies whose dividends per share are more than $1.00. Let \( \alpha = 0.05 \).
CALCULATING THE STATISTICS

11.48 Compute a one-way ANOVA on the following data. Use \( \alpha = .05 \). If there is a significant difference in treatment levels, use Tukey’s HSD to compute multiple comparisons. Let \( \alpha = .05 \) for the multiple comparisons.

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>7</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>9</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>6</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

11.49 Complete the following ANOVA table.

<table>
<thead>
<tr>
<th>SOURCE OF VARIANCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>249.61</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>317.80</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11.50 You are asked to analyze a completely randomized design that has six treatment levels and a total of 42 measurements. Complete the following table, which contains some information from the study.

<table>
<thead>
<tr>
<th>SOURCE OF VARIANCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>655</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11.51 Compute a one-way ANOVA on the following data. Let \( \alpha = .01 \). Use the Tukey-Kramer procedure to conduct multiple comparisons for the means.

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11.52 Examine the structure of the following experimental design. Determine which of the three designs presented in the chapter would be most likely to characterize this structure. Discuss the variables and the levels of variables. Determine the degrees of freedom.

11.53 Complete the following ANOVA table and determine whether there is any significance in treatment effects. Let \( \alpha = .05 \).

<table>
<thead>
<tr>
<th>SOURCE OF VARIANCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>20,994</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blocking</td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Error</td>
<td>33,891</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71,385</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11.54 Analyze the following data, gathered from a randomized block design using \( \alpha = .05 \). If there is a significant difference in the treatment effects, use Tukey’s HSD test to do multiple comparisons.

<table>
<thead>
<tr>
<th>TREATMENT</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>17</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>BLOCKING</td>
<td>3</td>
<td>20</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>VARIABLE</td>
<td>4</td>
<td>11</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>16</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>23</td>
<td>19</td>
<td>20</td>
</tr>
</tbody>
</table>

11.55 A two-way ANOVA has been computed on a factorial design. Treatment 1 has five levels and treatment 2 has two levels. There are four measures per cell. Complete the following ANOVA table. Use \( \alpha = .05 \) to test to determine significance of the effects. Comment on your findings.

<table>
<thead>
<tr>
<th>SOURCE OF VARIANCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>29.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment 2</td>
<td>12.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>73.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>110.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11.56 Compute a two-way ANOVA on the following data (\( \alpha = .01 \)).

<table>
<thead>
<tr>
<th>TREATMENT 1</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>TREATMENT 2</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>
### TESTING YOUR UNDERSTANDING

11.57 A company has conducted a consumer research project to ascertain customer service ratings from its customers. The customers were asked to rate the company on a scale from 1 to 7 on various quality characteristics. One question was the promptness of company response to a repair problem. The following data represent customer responses to this question. The customers were divided by geographic region and by age. Use analysis of variance to analyze the responses. Let $\alpha = .05$. Compute multiple comparisons where they are appropriate. Graph the cell means and observe any interaction.

<table>
<thead>
<tr>
<th>GEOGRAPHIC REGION</th>
<th>Southeast</th>
<th>West</th>
<th>Midwest</th>
<th>Northeast</th>
</tr>
</thead>
<tbody>
<tr>
<td>21–35</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>36–50</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Over 50</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

11.58 A major automobile manufacturer wants to know whether there is any difference in the average mileage of four different brands of tires (A, B, C, and D) because the manufacturer is trying to select the best supplier in terms of tire durability. The manufacturer selects comparable levels of tires from each company and tests some on comparable cars. The mileage results follow.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>31,000</td>
<td>24,000</td>
<td>30,500</td>
<td>24,500</td>
<td></td>
</tr>
<tr>
<td>25,000</td>
<td>25,500</td>
<td>28,000</td>
<td>27,000</td>
<td></td>
</tr>
<tr>
<td>28,500</td>
<td>27,000</td>
<td>32,500</td>
<td>26,000</td>
<td></td>
</tr>
<tr>
<td>29,000</td>
<td>26,500</td>
<td>28,000</td>
<td>21,000</td>
<td></td>
</tr>
<tr>
<td>32,000</td>
<td>25,000</td>
<td>31,000</td>
<td>25,500</td>
<td></td>
</tr>
<tr>
<td>27,500</td>
<td>28,000</td>
<td>26,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Use $\alpha = .05$ to test whether there is a significant difference in the mean mileage of these four brands. Assume tire mileage is normally distributed.

11.59 Agricultural researchers are studying three different ways of planting peanuts to determine whether significantly different levels of production yield will result. The researchers have access to a very large peanut farm on which to conduct their tests. They identify six blocks of land. In each block of land, peanuts are planted in each of the three different ways. At the end of the growing season, the peanuts are harvested and the average number of pounds per acre is determined for peanuts planted under each method in each block. Using the following data and $\alpha = .01$, test to determine whether there is a significant difference in yields among the planting methods.

<table>
<thead>
<tr>
<th>BLOCK</th>
<th>METHOD 1</th>
<th>METHOD 2</th>
<th>METHOD 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1310</td>
<td>1080</td>
<td>850</td>
</tr>
<tr>
<td>2</td>
<td>1275</td>
<td>1100</td>
<td>1020</td>
</tr>
<tr>
<td>3</td>
<td>1280</td>
<td>1050</td>
<td>780</td>
</tr>
<tr>
<td>4</td>
<td>1225</td>
<td>1020</td>
<td>870</td>
</tr>
<tr>
<td>5</td>
<td>1190</td>
<td>990</td>
<td>805</td>
</tr>
<tr>
<td>6</td>
<td>1300</td>
<td>1030</td>
<td>910</td>
</tr>
</tbody>
</table>

11.60 The Construction Labor Research Council lists a number of construction labor jobs that seem to pay approximately the same wages per hour. Some of these are bricklaying, iron working, and crane operation. Suppose a labor researcher takes a random sample of workers from each of these types of construction jobs and from across the country and asks what their hourly wages are. If this survey yields the following data, is there a significant difference in mean hourly wages for these three jobs? If there is a significant difference, use the Tukey-Kramer procedure to determine which pairs, if any, are also significantly different. Let $\alpha = .05$.

<table>
<thead>
<tr>
<th>JOB TYPE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bricklaying</td>
<td>19.25</td>
<td>24.33</td>
</tr>
<tr>
<td>Iron Working</td>
<td>20.50</td>
<td>22.90</td>
</tr>
<tr>
<td>Crane Operation</td>
<td>19.81</td>
<td>22.95</td>
</tr>
</tbody>
</table>

11.61 Why are mergers attractive to CEOs? One of the reasons might be a potential increase in market share that can come with the pooling of company markets. Suppose a random survey of CEOs is taken, and they are asked to respond on a scale from 1 to 5 (5 representing strongly agree) whether increase in market share is a good reason for considering a merger of their company with another. Suppose also that the data are as given here and that CEOs have been categorized by size of company and years they have been with their company. Use a two-way ANOVA to determine whether there are any significant differences in the responses to this question. Let $\alpha = .05$. 

<table>
<thead>
<tr>
<th></th>
<th>Bricklaying</th>
<th>Iron Working</th>
<th>Crane Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>22.90</td>
<td>21.20</td>
<td>22.95</td>
</tr>
<tr>
<td>21</td>
<td>23.30</td>
<td>25.52</td>
<td>22.90</td>
</tr>
<tr>
<td>22</td>
<td>21.00</td>
<td>25.55</td>
<td>16.20</td>
</tr>
</tbody>
</table>
### COMPANY SIZE
($ MILLION PER YEAR IN SALES)

<table>
<thead>
<tr>
<th></th>
<th>0-5</th>
<th>6-20</th>
<th>21-100</th>
<th>&gt;100</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0-2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>0-2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>0-2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>0-2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0-2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0-2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>0-2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>0-2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0-2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Over 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over 5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

11.62 Are some unskilled office jobs viewed as having more status than others? Suppose a study is conducted in which eight unskilled, unemployed people are interviewed. The people are asked to rate each of five positions on a scale from 1 to 10 to indicate the status of the position, with 10 denoting most status and 1 denoting least status. The resulting data are given here. Use $\alpha = .05$ to analyze these repeated measures randomized block design data.

<table>
<thead>
<tr>
<th>JOB</th>
<th>Mail Clerk</th>
<th>Typist</th>
<th>Receptionist</th>
<th>Secretary</th>
<th>Telephone Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Respondent</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Respondent</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Respondent</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Respondent</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Respondent</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Respondent</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Respondent</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

### INTERPRETING THE OUTPUT

11.63 Analyze the following MINITAB output. Describe the design of the experiment. Using $\alpha = .05$, determine whether there are any significant effects; if so, explain why. Discuss any other ramifications of the output.

#### One-Way Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>3</td>
<td>876.6</td>
<td>292.2</td>
<td>3.01</td>
<td>0.045</td>
</tr>
<tr>
<td>Error</td>
<td>32</td>
<td>3107.5</td>
<td>97.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>3984.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Individual 95% CIs for Mean Based on Pooled StDev

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8</td>
<td>307.73</td>
<td>5.98</td>
</tr>
<tr>
<td>C2</td>
<td>7</td>
<td>313.2</td>
<td>9.71</td>
</tr>
</tbody>
</table>

#### Analysis of Variance for depvar

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rowtreat</td>
<td>4</td>
<td>41.5</td>
<td>10.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coltreat</td>
<td>1</td>
<td>120.1</td>
<td>120.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>4</td>
<td>643.7</td>
<td>160.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>280.5</td>
<td>28.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>1085.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Excel output for ANOVA

11.64 Following is Excel output for an ANOVA problem. Describe the experimental design. The given value of alpha was .05. Discuss the output in terms of significant findings.

#### Two-Factor Without Replication

<table>
<thead>
<tr>
<th>SUMMARY</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>72</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>80</td>
<td>26.67</td>
<td>0.333</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>80</td>
<td>26.67</td>
<td>4.333</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>87</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>86</td>
<td>28.67</td>
<td>4.333</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>82</td>
<td>27.33</td>
<td>1.333</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>165</td>
<td>27.5</td>
<td>5.9</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>166</td>
<td>27.67</td>
<td>6.667</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>156</td>
<td>26</td>
<td>3.2</td>
</tr>
</tbody>
</table>

#### ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>48.278</td>
<td>5</td>
<td>9.656</td>
<td>3.16</td>
<td>0.057</td>
<td>3.33</td>
</tr>
<tr>
<td>Columns</td>
<td>10.111</td>
<td>2</td>
<td>5.056</td>
<td>1.65</td>
<td>0.239</td>
<td>4.10</td>
</tr>
<tr>
<td>Error</td>
<td>30.556</td>
<td>10</td>
<td>3.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>88.944</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11.65 Study the following MINITAB output and graph. Discuss the meaning of the output.

#### Two-way Analysis of Variance
### Anova: Two-Factor With Replication

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Count</strong></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>611</td>
<td>645</td>
<td>559</td>
<td>1815</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>203.67</td>
<td>215</td>
<td>186.33</td>
<td>201.67</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>6.333</td>
<td>1</td>
<td>2640.333</td>
<td>818.25</td>
</tr>
</tbody>
</table>

|        | 3        | 3        | 3        | 9     |
| **Sum**   | 657      | 681      | 698      | 2036  |
| **Average** | 219      | 227      | 232.67   | 226.22|
| **Variance** | 13       | 13       | 9.333    | 44.194|
| **Mean**  | 6.333    | 1        | 2640.333 | 818.25|
| **Total** | 30.0     | 32.5     | 35.0     | 37.5  |

|        | 4        | 3        | 3        | 9     |
| **Sum**   | 618      | 626      | 635      | 1879  |
| **Average** | 206      | 208.67   | 211.67   | 208.78|
| **Variance** | 9        | 6.333    | 2.333    | 10.444|

|        | 3        | 3        | 3        | 9     |
| **Sum**   | 628      | 631      | 629      | 1888  |
| **Average** | 209.33   | 210.33   | 209.67   | 209.78|
| **Variance** | 2.333    | 2.333    | 4.333    | 2.444 |

#### Source of Variation

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>2913.889</td>
<td>3</td>
<td>971.296</td>
<td>4.30</td>
<td>0.0146</td>
<td>3.01</td>
</tr>
<tr>
<td>Columns</td>
<td>240.389</td>
<td>2</td>
<td>120.194</td>
<td>0.53</td>
<td>0.5940</td>
<td>3.40</td>
</tr>
<tr>
<td>Interaction</td>
<td>1342.944</td>
<td>6</td>
<td>223.824</td>
<td>0.99</td>
<td>0.4533</td>
<td>2.51</td>
</tr>
<tr>
<td>Within</td>
<td>5419.333</td>
<td>24</td>
<td>225.806</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9916.556</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 11.68 Discuss the following MINITAB output.

#### One-Way Analysis of Variance

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>Mean</th>
<th>StDev</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>38.6</td>
<td></td>
<td>3.51</td>
<td>0.034</td>
</tr>
<tr>
<td>Error</td>
<td>262.2</td>
<td></td>
<td>13.1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>400.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Tukey's pairwise comparisons

- Individual error rate = 0.0500
- Individual error rate = 0.0111

#### Critical value = 3.96

<table>
<thead>
<tr>
<th>Intervals for (column level mean) - (row level mean)</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>-6.741</td>
<td>4.967</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-11.987</td>
<td>-11.100</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-9.369</td>
<td>2.339</td>
</tr>
</tbody>
</table>

#### 11.67 Study the following MINITAB output. Determine whether there are any significant effects and discuss the results. What kind of design was this and what was the size of it?

#### Two-way Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking</td>
<td>4</td>
<td>41.44</td>
<td>10.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>4</td>
<td>143.93</td>
<td>35.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>117.82</td>
<td>7.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>303.19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

12.53 Use the following data for parts (a) through (f).

<table>
<thead>
<tr>
<th>X</th>
<th>5</th>
<th>7</th>
<th>3</th>
<th>16</th>
<th>12</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>27</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>

a. Determine the equation of the least squares regression line to predict Y by X.
b. Using the X values, solve for the predicted values of Y and the residuals.
c. Solve for $S_e$.
d. Solve for $r^2$.
e. Test the slope of the regression line. Use $\alpha = .01$.
f. Comment on the results determined in parts (b)–(e), and make a statement about the fit of the line.

12.55 Solve for the value of $r$ for the following data.

<table>
<thead>
<tr>
<th>X</th>
<th>213</th>
<th>196</th>
<th>184</th>
<th>202</th>
<th>221</th>
<th>247</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>76</td>
<td>65</td>
<td>62</td>
<td>68</td>
<td>71</td>
<td>75</td>
</tr>
</tbody>
</table>

12.54 Use the following data for parts (a) through (h).

<table>
<thead>
<tr>
<th>X</th>
<th>53</th>
<th>47</th>
<th>41</th>
<th>50</th>
<th>58</th>
<th>62</th>
<th>45</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>11</td>
</tr>
</tbody>
</table>

a. Determine the equation of the simple regression line to predict Y from X.
b. Using the X values, solve for the predicted values of Y and the residuals.
c. Solve for SSE.
d. Calculate the standard error of the estimate.
e. Determine the coefficient of determination.
f. Calculate the coefficient of correlation.
g. Test the slope of the regression line. Assume $\alpha = .05$.
   What do you conclude about the slope?
h. Comment on parts (d) through (f).

If you were to develop a regression line to predict Y by X, what value would the coefficient of determination have?
12.56 Determine the equation of the least squares regression line to predict $Y$ from the following data.

<table>
<thead>
<tr>
<th>$X$</th>
<th>47</th>
<th>94</th>
<th>68</th>
<th>73</th>
<th>80</th>
<th>49</th>
<th>52</th>
<th>61</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>14</td>
<td>40</td>
<td>34</td>
<td>31</td>
<td>36</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
</tbody>
</table>

a. Construct a 95% confidence interval to estimate the mean $Y$ value for $X = 60$.
b. Construct a 95% prediction interval to estimate an individual $Y$ value for $X = 70$.
c. Interpret the results obtained in parts (a) and (b).

12.57 Determine the Pearson product–moment correlation coefficient for the following data.

<table>
<thead>
<tr>
<th>$X$</th>
<th>1</th>
<th>10</th>
<th>9</th>
<th>6</th>
<th>5</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

**TESTING YOUR UNDERSTANDING**

12.58 A manager of a car dealership believes there is a relationship between the number of salespeople on duty and the number of cars sold. Use the sample data collected for five different weeks at the dealership to calculate $r$. Is there much of a relationship? Solve for $r^2$. Explain what $r^2$ means in this problem.

<table>
<thead>
<tr>
<th>WEEK</th>
<th>NUMBER OF CARS SOLD</th>
<th>NUMBER OF SALESPEOPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>3</td>
</tr>
</tbody>
</table>

12.59 Executives of a video rental chain want to predict the success of a potential new store. The company’s researcher begins by gathering information on number of rentals and average family income from several of the chain’s present outlets.

<table>
<thead>
<tr>
<th>RENTALS</th>
<th>AVERAGE FAMILY INCOME ($1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>710</td>
<td>65</td>
</tr>
<tr>
<td>529</td>
<td>43</td>
</tr>
<tr>
<td>314</td>
<td>29</td>
</tr>
<tr>
<td>504</td>
<td>47</td>
</tr>
<tr>
<td>619</td>
<td>52</td>
</tr>
<tr>
<td>428</td>
<td>50</td>
</tr>
<tr>
<td>317</td>
<td>46</td>
</tr>
<tr>
<td>205</td>
<td>29</td>
</tr>
<tr>
<td>468</td>
<td>31</td>
</tr>
<tr>
<td>545</td>
<td>43</td>
</tr>
<tr>
<td>607</td>
<td>49</td>
</tr>
<tr>
<td>694</td>
<td>64</td>
</tr>
</tbody>
</table>

12.60 It seems logical that restaurant chains with more units (restaurants) would have greater sales. This is mitigated, however, by several possibilities: some units may be more profitable than others, some units may be larger, some units may serve more meals, some units may serve more expensive meals, and so on. The data shown here was published by Technomic, Inc. Use these data to determine whether there is a correlation between a restaurant chain’s sales and its number of units. How strong is the relationship?

<table>
<thead>
<tr>
<th>CHAIN</th>
<th>SALES ($ BILLIONS)</th>
<th>NUMBER OF UNITS (1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonald's</td>
<td>17.1</td>
<td>12.4</td>
</tr>
<tr>
<td>Burger King</td>
<td>7.9</td>
<td>7.5</td>
</tr>
<tr>
<td>Taco Bell</td>
<td>4.8</td>
<td>6.8</td>
</tr>
<tr>
<td>Pizza Hut</td>
<td>4.7</td>
<td>8.7</td>
</tr>
<tr>
<td>Wendy's</td>
<td>4.6</td>
<td>4.6</td>
</tr>
<tr>
<td>KFC</td>
<td>4.0</td>
<td>5.1</td>
</tr>
<tr>
<td>Subway</td>
<td>2.9</td>
<td>11.2</td>
</tr>
<tr>
<td>Dairy Queen</td>
<td>2.7</td>
<td>5.1</td>
</tr>
<tr>
<td>Hardee's</td>
<td>2.7</td>
<td>2.9</td>
</tr>
</tbody>
</table>

12.61 According to the National Marine Fisheries Service, the current landings in millions of pounds of fish by U.S. fleets are almost double what they were in the 1970s. In other words, fishing has not faded as an industry. However, the growth of this industry has varied by region as shown in the following data. Some regions have remained relatively constant, the South Atlantic region has dropped in pounds caught, and the Pacific-Alaska region has grown more than threefold.

<table>
<thead>
<tr>
<th>FISHERIES</th>
<th>1977</th>
<th>1997</th>
</tr>
</thead>
<tbody>
<tr>
<td>New England</td>
<td>581</td>
<td>642</td>
</tr>
<tr>
<td>Mid-Atlantic</td>
<td>213</td>
<td>242</td>
</tr>
<tr>
<td>Chesapeake</td>
<td>668</td>
<td>729</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>345</td>
<td>269</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td>1476</td>
<td>1497</td>
</tr>
<tr>
<td>Pacific-Alaska</td>
<td>1776</td>
<td>6129</td>
</tr>
</tbody>
</table>

a. Compute the correlation coefficient between the 1977 and 1997 data. What does this tell, if anything, about the industry?
b. Develop a simple regression model to predict the 1997 landings by the 1977 landings. According to the model, if a region had 700 landings in 1977, what would the predicted number be for 1997? Construct a confidence interval for the average $Y$ value for the 700 landings. Use the $t$ statistic to test to determine whether the slope is significantly different from zero. Use $\alpha = .05$.

12.62 People in the aerospace industry believe the cost of a space project is a function of the weight of the major object being sent into space. Use the following data to develop a regression model to predict the cost of a space project as a function of the weight of the object.
12.63 The following data represent a breakdown of state banks and all savings organizations in the United States according to the Federal Reserve System.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>STATE BANKS</th>
<th>ALL SAVINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>1342</td>
<td>2330</td>
</tr>
<tr>
<td>1945</td>
<td>1864</td>
<td>2667</td>
</tr>
<tr>
<td>1950</td>
<td>1912</td>
<td>3054</td>
</tr>
<tr>
<td>1955</td>
<td>1847</td>
<td>3764</td>
</tr>
<tr>
<td>1960</td>
<td>1641</td>
<td>4423</td>
</tr>
<tr>
<td>1965</td>
<td>1405</td>
<td>4837</td>
</tr>
<tr>
<td>1970</td>
<td>1147</td>
<td>4694</td>
</tr>
<tr>
<td>1975</td>
<td>1046</td>
<td>4407</td>
</tr>
<tr>
<td>1980</td>
<td>997</td>
<td>4328</td>
</tr>
<tr>
<td>1985</td>
<td>1070</td>
<td>3626</td>
</tr>
<tr>
<td>1990</td>
<td>1099</td>
<td>2815</td>
</tr>
<tr>
<td>1995</td>
<td>1042</td>
<td>2030</td>
</tr>
<tr>
<td>1997</td>
<td>992</td>
<td>1779</td>
</tr>
</tbody>
</table>

a. Develop a regression model to predict the total number of state banks by the number of all savings organizations.

b. Determine the correlation between the number of state banks and the number of all savings organizations.

12.64 How strong is the correlation between the inflation rate and 30-year treasury yields? The following data published by Fuji Securities, Inc., are given as pairs of inflation rates and treasury yields for selected years over a 35-year period.

<table>
<thead>
<tr>
<th>INFLATION RATE</th>
<th>30-YEAR TREASURY YIELD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.57%</td>
<td>3.05%</td>
</tr>
<tr>
<td>2.23</td>
<td>3.93</td>
</tr>
<tr>
<td>2.17</td>
<td>4.68</td>
</tr>
<tr>
<td>4.53</td>
<td>6.57</td>
</tr>
<tr>
<td>7.25</td>
<td>8.27</td>
</tr>
<tr>
<td>9.25</td>
<td>12.01</td>
</tr>
<tr>
<td>5.00</td>
<td>10.27</td>
</tr>
<tr>
<td>4.62</td>
<td>8.45</td>
</tr>
</tbody>
</table>

Compute the Pearson product-moment correlation coefficient to determine the strength of the correlation between these two variables. Comment on the strength and direction of the correlation.

12.65 Is the amount of money spent by companies on advertising a function of the total sales of the company? Shown are sales income and advertising cost data for seven companies, published by Advertising Age.

<table>
<thead>
<tr>
<th>COMPANY</th>
<th>ADVERTISING (MILLIONS)</th>
<th>SALES (BILLIONS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procter &amp; Gamble</td>
<td>$1703.1</td>
<td>37.1</td>
</tr>
<tr>
<td>Philip Morris</td>
<td>1319.0</td>
<td>56.1</td>
</tr>
<tr>
<td>Ford Motor</td>
<td>973.1</td>
<td>153.6</td>
</tr>
<tr>
<td>PepsiCo</td>
<td>797.4</td>
<td>20.9</td>
</tr>
<tr>
<td>Time Warner</td>
<td>779.1</td>
<td>13.3</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>738.7</td>
<td>22.6</td>
</tr>
<tr>
<td>MCI</td>
<td>455.4</td>
<td>19.7</td>
</tr>
</tbody>
</table>

Use the data to develop a regression line to predict the amount of advertising by sales. Compute $S_e$ and $r^2$. Assuming $\alpha = .05$, test the slope of the regression line. Comment on the strength of the regression model.

12.66 Can the consumption of water in a city be predicted by temperature? The following data represent a sample of a day’s water consumption and the high temperature for that day.

<table>
<thead>
<tr>
<th>WATER USE (MILLION GAL)</th>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>219</td>
<td>103*</td>
</tr>
<tr>
<td>56</td>
<td>39</td>
</tr>
<tr>
<td>107</td>
<td>77</td>
</tr>
<tr>
<td>129</td>
<td>78</td>
</tr>
<tr>
<td>68</td>
<td>50</td>
</tr>
<tr>
<td>184</td>
<td>96</td>
</tr>
<tr>
<td>150</td>
<td>90</td>
</tr>
<tr>
<td>112</td>
<td>75</td>
</tr>
</tbody>
</table>

Develop a least squares regression line to predict the amount of water used in a day in a city by the high temperature for that day. What would be the predicted water usage for a temperature of 100°? Evaluate the regression model by calculating $S_e$ by calculating $r^2$, and by testing the slope. Let $\alpha = .01$.

INTERPRETING THE OUTPUT

12.67 Study the following MINITAB output from a regression analysis to predict $Y$ from $X$.

a. What is the equation of the regression model?

b. What is the meaning of the coefficient of $X$?

c. What is the result of the test of the slope of the regression model? Let $\alpha = .10$. Why is the $t$ ratio negative?

d. Comment on $r^2$ and the standard error of the estimate.

e. Comment on the relationship of the $F$ value to the $t$ ratio for $X$.

f. The correlation coefficient for these two variables is $- .7918$. Is this surprising to you? Why or why not?
Regression Analysis

The regression equation is
\[ Y = 67.2 - 0.0565 \times X \]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Predictor} & \text{Coef} & \text{StDev} & \text{T} & \text{p} \\
\hline
\text{Constant} & 67.231 & 5.046 & 13.32 & 0.000 \\
\text{X} & -0.05650 & 0.01027 & -5.50 & 0.000 \\
\hline
\end{array}
\]

\[ S = 10.32 \quad \text{R-Sq} = 62.7\% \quad \text{R-Sq(adj)} = 60.6\% \]

Analysis of Variance

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Source} & \text{DF} & \text{SS} & \text{MS} & \text{F} & \text{p} \\
\hline
\text{Regression} & 1 & 3222.9 & 3222.9 & 30.25 & 0.000 \\
\text{Residual Error} & 18 & 1918.0 & 106.6 & \\
\text{Total} & 19 & 5141.0 & & \\
\hline
\end{array}
\]

12.68 Study the following MINITAB regression output for a model to predict the daily room rate of lodging in a given city by the daily car-rental rate in that city.

Regression Analysis

The regression equation is
\[ \text{RoomRate} = 80.5 + 0.960 \times \text{CarRate} \]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Predictor} & \text{Coef} & \text{StDev} & \text{T} & \text{p} \\
\hline
\text{Constant} & 80.49 & 33.03 & 2.44 & 0.025 \\
\text{CarRate} & 0.9601 & 0.6831 & 1.41 & 0.177 \\
\hline
\end{array}
\]

\[ S = 28.29 \quad \text{R-sq} = 9.9\% \quad \text{R-sq(adj)} = 4.9\% \]

Analysis of Variance

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Source} & \text{DF} & \text{SS} & \text{MS} & \text{F} & \text{p} \\
\hline
\text{Regression} & 1 & 1580.9 & 1580.9 & 1.98 & 0.177 \\
\text{Error} & 18 & 14406.1 & 800.3 & \\
\text{Total} & 19 & 15987.0 & & \\
\hline
\end{array}
\]

a. What is the equation of the regression model? What does it mean in terms of daily room rate and daily car-rental rate?
b. What is the value of $r^2$? Is it high or low?
c. What is the standard error of the estimate? In light of the data being analyzed, what is the meaning of $S^2$?
d. Discuss the outcome of the hypothesis test of the slope and the $F$ test for overall predictability of the model.
e. If the daily car-rental rate in a city is $45$, what is the predicted daily room rate according to the regression model?
f. In general, is this a strong model?
12.70 Study the following Excel regression output for an analysis attempting to predict the number of union members in the United States by the size of the labor force for selected years over a 30-year period from data published by the U.S. Bureau of Labor Statistics. Analyze the computer output. Discuss the strength of the model in terms of proportion of variation accounted for, slope, and overall predictability. Using the equation of the regression line, attempt to predict the number of union members when the labor force is 100,000. Note that the model was developed with data already recoded in 1000 units. Use the data in the model as is.

**SUMMARY OUTPUT**

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>22348.97</td>
<td>1846.367</td>
<td>12.10</td>
<td>4.4E-08</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>-0.0524</td>
<td>0.0196</td>
<td>-2.67</td>
<td>0.0205</td>
</tr>
</tbody>
</table>

**RESIDUAL OUTPUT**

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted Y</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19161.39</td>
<td>-1862.39</td>
</tr>
<tr>
<td>2</td>
<td>18631.75</td>
<td>749.25</td>
</tr>
<tr>
<td>3</td>
<td>18315.95</td>
<td>1295.05</td>
</tr>
<tr>
<td>4</td>
<td>17602.12</td>
<td>2240.88</td>
</tr>
<tr>
<td>5</td>
<td>17516.68</td>
<td>-176.68</td>
</tr>
<tr>
<td>6</td>
<td>17394.71</td>
<td>-398.71</td>
</tr>
<tr>
<td>7</td>
<td>17269.86</td>
<td>-294.86</td>
</tr>
<tr>
<td>8</td>
<td>17144.07</td>
<td>-231.07</td>
</tr>
<tr>
<td>9</td>
<td>17033.79</td>
<td>-31.79</td>
</tr>
<tr>
<td>10</td>
<td>16925.13</td>
<td>34.87</td>
</tr>
<tr>
<td>11</td>
<td>16902.86</td>
<td>-162.86</td>
</tr>
<tr>
<td>12</td>
<td>16961.51</td>
<td>-393.51</td>
</tr>
<tr>
<td>13</td>
<td>16914.23</td>
<td>-524.23</td>
</tr>
<tr>
<td>14</td>
<td>16841.95</td>
<td>-243.95</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>6868285.79</td>
<td>6868286</td>
<td>7.12</td>
<td>0.00205</td>
</tr>
<tr>
<td>Residual</td>
<td>12</td>
<td>11577055.64</td>
<td>964755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>18445341.43</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ANALYZING THE DATABASES**

1. Use the manufacturing database to correlate the Number of Employees with the Number of Production Workers. Is this a high correlation? What would you expect the correlation to be? Is there a strong correlation between Cost of Materials and Value Added by Manufacture? Why do you think it is this way?

2. Develop a regression model from the manufacturing database to predict New Capital Expenditures from Value Added by Manufacture. Discuss the model and its strength on the basis of indicators presented in this chapter. Does it seem logical that the dollars spent on New Capital Expenditure could be predicted by Value Added by Manufacture?

3. Using the hospital database, develop a regression model to predict the number of Personnel by the number of Births. Now develop a regression model to predict number of Personnel by number of Beds. Examine the regression output. Which model is stronger in predicting number of Personnel? Explain why, using techniques presented in this chapter. Use the second regression model to predict the number of Personnel in a hospital that has 110 beds. Construct a 95% confidence interval around this prediction for the average value of $Y$.

4. Produce a correlation matrix for the variables Beds, Admissions, Census, Outpatient Visits, Births, Total Expenditures, Payroll Expenditures, and Personnel for the hospital database. Which variables are most highly correlated? Which variables are least correlated?

5. Analyze all the variables except Type in the financial database by using a correlation matrix. The seven variables in this database are capable of producing 21 pairs of correlations. Which are most highly correlated? Select the variable that is most highly correlated with P/E ratio and use it as a predictor to develop a regression model to predict P/E ratio. How did the model do?

6. Use the stock market database to develop a regression model to predict the Utility Index by the Stock Volume. How well did the model perform? Did it perform as you expected? Why or why not? Construct a correlation matrix for the variables of this database (excluding Part of Month) so that you can explore the stock market. Did you discover any apparent relationships between variables?
CALCULATING THE STATISTICS

13.40 Use the following data to develop a multiple regression model to predict \( Y \) from \( X_1 \) and \( X_2 \). Discuss the output, including comments about the overall strength of the model, the significance of the regression coefficients, and other indicators of model fit.

\[
\begin{array}{cccc}
X_1 & X_2 & Y \\
29 & 1.64 & 198 \\
71 & 2.81 & 214 \\
54 & 2.22 & 211 \\
73 & 2.70 & 219 \\
67 & 1.57 & 184 \\
32 & 1.63 & 167 \\
47 & 1.99 & 201 \\
43 & 2.14 & 204 \\
60 & 2.04 & 190 \\
32 & 2.93 & 222 \\
34 & 2.15 & 197 \\
\end{array}
\]

13.41 Given here are the data for a dependent variable, \( Y \), and independent variables. Use these data to develop a regression model to predict \( Y \). Discuss the output. Which variable is an indicator variable? Was it a significant predictor of \( Y \)?

\[
\begin{array}{cccc}
X_1 & X_2 & X_3 & Y \\
0 & 51 & 16.4 & 14 \\
0 & 48 & 17.1 & 17 \\
1 & 29 & 18.2 & 29 \\
0 & 36 & 17.9 & 32 \\
0 & 40 & 16.5 & 54 \\
1 & 27 & 17.1 & 86 \\
1 & 14 & 17.8 & 117 \\
0 & 17 & 18.2 & 120 \\
1 & 16 & 16.9 & 194 \\
1 & 9 & 18.0 & 203 \\
1 & 14 & 18.9 & 217 \\
0 & 11 & 18.5 & 235 \\
\end{array}
\]
13.42 Use the following data and a stepwise regression analysis to predict \( Y \). In addition to the two independent variables given here, include three other predictors in your analysis: the square of each \( X \) as a predictor and an interaction predictor. Discuss the results of the process.

\[
\begin{array}{ccc}
X_1 & X_2 & Y \\
10 & 3 & 2002 \\
5 & 14 & 1747 \\
8 & 4 & 1980 \\
7 & 4 & 1902 \\
6 & 7 & 1842 \\
6 & 6 & 1883 \\
4 & 21 & 1697 \\
11 & 4 & 2021 \\
\end{array}
\]

\[
\begin{array}{ccc}
X_1 & X_2 & Y \\
5 & 12 & 1750 \\
6 & 8 & 1832 \\
5 & 18 & 1795 \\
7 & 4 & 1917 \\
8 & 5 & 1943 \\
6 & 9 & 1830 \\
5 & 12 & 1786 \\
\end{array}
\]

13.43 Use the \( X_1 \) values and the log of the \( X_1 \) values given here to predict the \( Y \) values by using a stepwise regression procedure. Discuss the output. Were either or both of the predictors significant?

\[
\begin{array}{ccc}
Y & X_1 & Y \\
20.4 & 850 & 13.2 \\
11.6 & 146 & 17.5 \\
17.8 & 521 & 12.4 \\
15.3 & 304 & 10.6 \\
22.4 & 1029 & 19.8 \\
21.9 & 910 & 17.4 \\
16.4 & 242 & 19.4 \\
\end{array}
\]

13.45 The U.S. Bureau of Mines produces data on the price of minerals. Shown here are the average prices per year for several minerals over a decade. Use these data and a stepwise regression procedure to produce a model to predict the average price of gold from the other variables. Comment on the results of the process.

<table>
<thead>
<tr>
<th>MINERAL</th>
<th>PRICE ($) PER OZ</th>
<th>PRICE ($ PER LB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOLD</td>
<td>161.1</td>
<td>64.2</td>
</tr>
<tr>
<td>COPPER</td>
<td>308.0</td>
<td>93.3</td>
</tr>
<tr>
<td>SILVER</td>
<td>613.0</td>
<td>101.3</td>
</tr>
<tr>
<td>ALUMINUM</td>
<td>460.0</td>
<td>84.2</td>
</tr>
<tr>
<td></td>
<td>376.0</td>
<td>72.8</td>
</tr>
<tr>
<td></td>
<td>424.0</td>
<td>76.5</td>
</tr>
<tr>
<td></td>
<td>361.0</td>
<td>66.8</td>
</tr>
<tr>
<td></td>
<td>318.0</td>
<td>67.0</td>
</tr>
<tr>
<td></td>
<td>368.0</td>
<td>66.1</td>
</tr>
<tr>
<td></td>
<td>448.0</td>
<td>82.5</td>
</tr>
<tr>
<td></td>
<td>438.0</td>
<td>120.5</td>
</tr>
<tr>
<td></td>
<td>382.6</td>
<td>130.9</td>
</tr>
</tbody>
</table>

13.46 The Shipbuilders Council of America in Washington, DC, publishes data about private shipyards. Among the variables reported by this organization are the employment figures (per 1000), the number of naval vessels under construction, and the number of repairs or conversions done to commercial ships (in millions of dollars). Shown here are the data for these three variables over a 7-year period. Use the data to develop a regression model to predict private shipyard employment from number of naval vessels under construction and repairs or conversions of commercial ships. Graph each of these predictors separately with the response variable and use Tukey’s four-quadrant approach to explore possible recoding schemes for nonlinear relationships. Include any of these in the regression model. Comment on the regression model and its strengths and/or its weaknesses.

13.44 The U.S. Commodities Futures Trading Commission reports on the volume of trading in the U.S. commodity futures exchanges. Shown here are the figures for grain, oilseeds, and livestock products over a period of several years. Use these data to develop a multiple regression model to predict grain futures volume of trading from oilseeds volume and livestock products volume. All figures are given in units of millions. Graph each of these predictors separately with the response variable and use Tukey’s four-quadrant approach to explore possible recoding schemes for nonlinear relationships. Include any of these in the regression model. Comment on the results.

<table>
<thead>
<tr>
<th>GRAIN</th>
<th>OILSEEDS</th>
<th>LIVESTOCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>3.7</td>
<td>3.4</td>
</tr>
<tr>
<td>18.3</td>
<td>15.7</td>
<td>11.8</td>
</tr>
<tr>
<td>19.8</td>
<td>20.3</td>
<td>9.8</td>
</tr>
<tr>
<td>14.9</td>
<td>15.8</td>
<td>11.0</td>
</tr>
<tr>
<td>17.8</td>
<td>19.8</td>
<td>11.1</td>
</tr>
<tr>
<td>15.9</td>
<td>23.5</td>
<td>8.4</td>
</tr>
<tr>
<td>10.7</td>
<td>14.9</td>
<td>7.9</td>
</tr>
<tr>
<td>13.8</td>
<td>14.2</td>
<td>8.8</td>
</tr>
<tr>
<td>10.9</td>
<td>14.2</td>
<td>9.6</td>
</tr>
<tr>
<td>15.9</td>
<td>22.5</td>
<td>8.2</td>
</tr>
</tbody>
</table>
Develop any other appropriate predictor variables by recoding data and include them in the analysis. Comment on the result of this analysis.

### ALL COMMODITIES

<table>
<thead>
<tr>
<th>FOOD</th>
<th>SHELTER</th>
<th>APPAREL</th>
<th>FUEL OIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>.9</td>
<td>1.0</td>
<td>2.0</td>
<td>1.6</td>
</tr>
<tr>
<td>.6</td>
<td>1.3</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td>.9</td>
<td>1.6</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>1.2</td>
<td>1.3</td>
<td>1.5</td>
<td>.9</td>
</tr>
<tr>
<td>1.1</td>
<td>2.2</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>2.6</td>
<td>5.0</td>
<td>3.0</td>
<td>2.5</td>
</tr>
<tr>
<td>1.9</td>
<td>.9</td>
<td>3.6</td>
<td>4.1</td>
</tr>
<tr>
<td>3.5</td>
<td>3.5</td>
<td>4.5</td>
<td>5.3</td>
</tr>
<tr>
<td>4.7</td>
<td>5.1</td>
<td>8.3</td>
<td>5.8</td>
</tr>
<tr>
<td>4.5</td>
<td>5.7</td>
<td>8.9</td>
<td>4.2</td>
</tr>
<tr>
<td>3.6</td>
<td>3.1</td>
<td>4.2</td>
<td>3.2</td>
</tr>
<tr>
<td>3.0</td>
<td>4.2</td>
<td>4.6</td>
<td>2.0</td>
</tr>
<tr>
<td>7.4</td>
<td>14.5</td>
<td>4.7</td>
<td>3.7</td>
</tr>
<tr>
<td>11.9</td>
<td>14.3</td>
<td>9.6</td>
<td>7.4</td>
</tr>
<tr>
<td>8.8</td>
<td>8.5</td>
<td>9.9</td>
<td>4.5</td>
</tr>
<tr>
<td>4.3</td>
<td>3.0</td>
<td>5.5</td>
<td>3.7</td>
</tr>
<tr>
<td>5.8</td>
<td>6.3</td>
<td>6.6</td>
<td>4.5</td>
</tr>
<tr>
<td>7.2</td>
<td>9.9</td>
<td>10.2</td>
<td>3.6</td>
</tr>
<tr>
<td>11.3</td>
<td>11.0</td>
<td>13.9</td>
<td>4.3</td>
</tr>
</tbody>
</table>

13.48 The U.S. Department of Agriculture publishes data annually on various selected farm products. Shown here are the unit production figures for three farm products for 10 years during a 20-year period. Use these data and a stepwise regression analysis to predict corn production by the production of soybeans and wheat. Comment on the results.

<table>
<thead>
<tr>
<th>CORN (MILLION BUSHELS)</th>
<th>SOYBEANS (MILLION BUSHELS)</th>
<th>WHEAT (MILLION BUSHELS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4152</td>
<td>1127</td>
<td>1352</td>
</tr>
<tr>
<td>6639</td>
<td>1798</td>
<td>2381</td>
</tr>
<tr>
<td>4175</td>
<td>1636</td>
<td>2420</td>
</tr>
<tr>
<td>7672</td>
<td>1861</td>
<td>2595</td>
</tr>
<tr>
<td>8876</td>
<td>2099</td>
<td>2424</td>
</tr>
<tr>
<td>8226</td>
<td>1940</td>
<td>2091</td>
</tr>
<tr>
<td>7131</td>
<td>1938</td>
<td>2108</td>
</tr>
<tr>
<td>4929</td>
<td>1549</td>
<td>1812</td>
</tr>
<tr>
<td>7525</td>
<td>1924</td>
<td>2037</td>
</tr>
<tr>
<td>7933</td>
<td>1922</td>
<td>2739</td>
</tr>
</tbody>
</table>

13.49 Cost-of-living indexes for selected metropolitan areas have been accumulated by the American Chamber of Commerce Researchers Association. Shown here are cost-of-living indexes for 25 different cities on five different items for a recent year. Use the data to develop a regression model to predict the grocery cost-of-living index by the indexes of housing, utilities, transportation, and healthcare. Discuss the results, highlighting both the significant and nonsignificant predictors.

<table>
<thead>
<tr>
<th>CITY</th>
<th>GROCERY ITEMS</th>
<th>HOUSING</th>
<th>UTILITIES</th>
<th>TRANSPORTATION</th>
<th>HEALTHCARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albany</td>
<td>108.3</td>
<td>106.8</td>
<td>127.4</td>
<td>89.1</td>
<td>107.5</td>
</tr>
<tr>
<td>Albuquerque</td>
<td>96.3</td>
<td>105.2</td>
<td>98.8</td>
<td>100.9</td>
<td>102.1</td>
</tr>
<tr>
<td>Augusta, GA</td>
<td>96.2</td>
<td>88.8</td>
<td>115.6</td>
<td>102.3</td>
<td>94.0</td>
</tr>
<tr>
<td>Austin</td>
<td>98.0</td>
<td>83.9</td>
<td>87.7</td>
<td>97.4</td>
<td>94.9</td>
</tr>
<tr>
<td>Baltimore</td>
<td>106.0</td>
<td>114.1</td>
<td>108.1</td>
<td>112.8</td>
<td>111.5</td>
</tr>
<tr>
<td>Buffalo</td>
<td>103.1</td>
<td>117.3</td>
<td>127.6</td>
<td>107.8</td>
<td>100.8</td>
</tr>
<tr>
<td>Colorado Springs</td>
<td>94.5</td>
<td>88.5</td>
<td>74.6</td>
<td>93.3</td>
<td>102.4</td>
</tr>
<tr>
<td>Dallas</td>
<td>105.4</td>
<td>98.9</td>
<td>108.9</td>
<td>110.0</td>
<td>106.8</td>
</tr>
<tr>
<td>Denver</td>
<td>91.5</td>
<td>108.3</td>
<td>97.2</td>
<td>105.9</td>
<td>114.3</td>
</tr>
<tr>
<td>Des Moines</td>
<td>94.3</td>
<td>95.1</td>
<td>111.4</td>
<td>105.7</td>
<td>96.2</td>
</tr>
<tr>
<td>El Paso</td>
<td>102.9</td>
<td>94.6</td>
<td>90.9</td>
<td>104.2</td>
<td>91.4</td>
</tr>
<tr>
<td>Indianapolis</td>
<td>96.0</td>
<td>99.7</td>
<td>92.1</td>
<td>102.7</td>
<td>97.4</td>
</tr>
<tr>
<td>Jacksonville, FL</td>
<td>96.1</td>
<td>90.4</td>
<td>96.0</td>
<td>106.0</td>
<td>96.1</td>
</tr>
<tr>
<td>Kansas City</td>
<td>89.8</td>
<td>92.4</td>
<td>96.3</td>
<td>95.6</td>
<td>93.6</td>
</tr>
<tr>
<td>Knoxville</td>
<td>93.2</td>
<td>88.0</td>
<td>91.7</td>
<td>91.6</td>
<td>82.3</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>103.3</td>
<td>211.3</td>
<td>75.6</td>
<td>102.1</td>
<td>128.5</td>
</tr>
<tr>
<td>Louisville</td>
<td>94.6</td>
<td>91.0</td>
<td>79.4</td>
<td>102.4</td>
<td>88.4</td>
</tr>
<tr>
<td>Memphis</td>
<td>99.1</td>
<td>86.2</td>
<td>91.1</td>
<td>101.1</td>
<td>85.5</td>
</tr>
<tr>
<td>Miami</td>
<td>100.3</td>
<td>123.0</td>
<td>125.6</td>
<td>104.3</td>
<td>137.8</td>
</tr>
<tr>
<td>Minneapolis</td>
<td>92.8</td>
<td>112.3</td>
<td>105.2</td>
<td>106.0</td>
<td>107.5</td>
</tr>
<tr>
<td>Mobile</td>
<td>99.9</td>
<td>81.1</td>
<td>104.9</td>
<td>102.8</td>
<td>92.2</td>
</tr>
<tr>
<td>Nashville</td>
<td>95.8</td>
<td>107.7</td>
<td>91.6</td>
<td>98.1</td>
<td>90.9</td>
</tr>
<tr>
<td>New Orleans</td>
<td>104.0</td>
<td>83.4</td>
<td>122.2</td>
<td>98.2</td>
<td>87.0</td>
</tr>
<tr>
<td>Oklahoma City</td>
<td>98.2</td>
<td>79.4</td>
<td>103.4</td>
<td>97.3</td>
<td>97.1</td>
</tr>
<tr>
<td>Phoenix</td>
<td>95.7</td>
<td>98.7</td>
<td>96.3</td>
<td>104.6</td>
<td>115.2</td>
</tr>
</tbody>
</table>
13.50 Work Problem 11.11 in Chapter 11 by using multiple regression techniques. Compare and contrast the regression results with the ANOVA results.

13.51 Work Problem 11.13 in Chapter 11 by using regression techniques. Discuss the results.

13.52 Work Problem 11.51 from Chapter 11 by using techniques learned in this chapter. Compare your results with the results obtained by using standard ANOVA techniques.

**INTERPRETING THE OUTPUT**

13.53 A stepwise regression procedure was used to analyze a set of 20 observations taken on four predictor variables to predict a dependent variable. The results of this procedure are given next. Discuss the results.

**STEPWISE REGRESSION OF Y ON 4 PREDICTORS, WITH N = 20**

<table>
<thead>
<tr>
<th>STEP</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>152.2</td>
<td>124.5</td>
</tr>
<tr>
<td>$X_1$</td>
<td>-50.6</td>
<td>-43.4</td>
</tr>
<tr>
<td>T-RATIO</td>
<td>7.42</td>
<td>6.13</td>
</tr>
<tr>
<td>$X_2$</td>
<td>1.36</td>
<td>2.13</td>
</tr>
<tr>
<td>T-RATIO</td>
<td>15.2</td>
<td>13.9</td>
</tr>
<tr>
<td>R-SQ</td>
<td>75.39</td>
<td>80.59</td>
</tr>
</tbody>
</table>

13.54 Shown here are the data for $Y$ and three predictors, $X_1$, $X_2$, and $X_3$. A stepwise regression procedure has been done on these data; the results are also given. Comment on the outcome of the stepwise analysis in light of the data.

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>21</td>
<td>1</td>
<td>204</td>
</tr>
<tr>
<td>97</td>
<td>25</td>
<td>0</td>
<td>198</td>
</tr>
<tr>
<td>93</td>
<td>22</td>
<td>1</td>
<td>184</td>
</tr>
<tr>
<td>95</td>
<td>27</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>90</td>
<td>29</td>
<td>1</td>
<td>182</td>
</tr>
<tr>
<td>91</td>
<td>20</td>
<td>1</td>
<td>159</td>
</tr>
<tr>
<td>91</td>
<td>18</td>
<td>1</td>
<td>147</td>
</tr>
<tr>
<td>94</td>
<td>25</td>
<td>0</td>
<td>196</td>
</tr>
<tr>
<td>98</td>
<td>26</td>
<td>0</td>
<td>228</td>
</tr>
<tr>
<td>99</td>
<td>24</td>
<td>0</td>
<td>242</td>
</tr>
<tr>
<td>90</td>
<td>28</td>
<td>1</td>
<td>162</td>
</tr>
<tr>
<td>92</td>
<td>23</td>
<td>1</td>
<td>180</td>
</tr>
<tr>
<td>96</td>
<td>25</td>
<td>0</td>
<td>219</td>
</tr>
</tbody>
</table>

**Regression Statistics**

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th>Multiplke R</th>
<th>0.567</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Square</td>
<td>0.321</td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.208</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>159.681</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>289856.08</td>
<td>72464.02</td>
<td>2.84</td>
</tr>
<tr>
<td>Residual</td>
<td>24</td>
<td>611955.23</td>
<td>25498.13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>901811.31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coefficients**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>336.79</td>
<td>124.08</td>
<td>2.71</td>
</tr>
<tr>
<td>$X_1$</td>
<td>1.65</td>
<td>1.78</td>
<td>0.93</td>
</tr>
<tr>
<td>$X_2$</td>
<td>-5.63</td>
<td>13.47</td>
<td>-0.42</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.26</td>
<td>1.68</td>
<td>0.16</td>
</tr>
<tr>
<td>$X_4$</td>
<td>185.50</td>
<td>66.22</td>
<td>2.80</td>
</tr>
</tbody>
</table>

**Regression Statistics**

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th>Multiplke R</th>
<th>0.566</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Square</td>
<td>0.321</td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>156.534</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>289238.1</td>
<td>96412.7</td>
<td>3.93</td>
</tr>
<tr>
<td>Residual</td>
<td>25</td>
<td>612573.2</td>
<td>24502.9</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>901811.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coefficients**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>342.919</td>
<td>115.34</td>
<td>2.97</td>
</tr>
<tr>
<td>$X_1$</td>
<td>1.834</td>
<td>1.31</td>
<td>1.40</td>
</tr>
<tr>
<td>$X_2$</td>
<td>-5.749</td>
<td>13.18</td>
<td>-0.44</td>
</tr>
<tr>
<td>$X_4$</td>
<td>181.220</td>
<td>59.05</td>
<td>3.07</td>
</tr>
</tbody>
</table>
**SUPPLEMENTARY PROBLEMS**

**CALCULATING THE STATISTICS**

14.14 Compute unweighted aggregate price index numbers for each of the years given here using 1975 as the base year.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.70</td>
<td>.72</td>
<td>.79</td>
<td>.86</td>
<td>.84</td>
</tr>
<tr>
<td>2</td>
<td>1.71</td>
<td>1.65</td>
<td>1.72</td>
<td>1.84</td>
<td>1.92</td>
</tr>
<tr>
<td>3</td>
<td>.12</td>
<td>.15</td>
<td>.21</td>
<td>.30</td>
<td>.26</td>
</tr>
<tr>
<td>4</td>
<td>1.14</td>
<td>1.32</td>
<td>1.51</td>
<td>1.48</td>
<td>1.43</td>
</tr>
</tbody>
</table>

14.15 Using the following data and 1996 as the base year, compute the Laspeyres price index for 1998, the Paasche price index for 1997, and Fisher's ideal price index for 1999.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$2.75</td>
<td>12</td>
<td>$2.98</td>
<td>9</td>
<td>$3.10</td>
<td>9</td>
<td>$3.21</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>0.85</td>
<td>47</td>
<td>0.89</td>
<td>52</td>
<td>0.95</td>
<td>61</td>
<td>0.98</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>1.33</td>
<td>20</td>
<td>1.32</td>
<td>28</td>
<td>1.36</td>
<td>25</td>
<td>1.40</td>
<td>32</td>
</tr>
</tbody>
</table>

14.16 Compute index numbers for the following data using 1985 as the base year.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1.28</td>
<td>138</td>
<td>$1.75</td>
<td>105</td>
</tr>
<tr>
<td>2</td>
<td>.90</td>
<td>255</td>
<td>.96</td>
<td>240</td>
</tr>
<tr>
<td>3</td>
<td>2.24</td>
<td>30</td>
<td>2.19</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>.17</td>
<td>970</td>
<td>.29</td>
<td>1024</td>
</tr>
</tbody>
</table>

14.17 Use the following data to answer the questions. Assume that 1980 is the base year.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1.28</td>
<td>138</td>
<td>$1.75</td>
<td>105</td>
</tr>
<tr>
<td>2</td>
<td>.90</td>
<td>255</td>
<td>.96</td>
<td>240</td>
</tr>
<tr>
<td>3</td>
<td>2.24</td>
<td>30</td>
<td>2.19</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>.17</td>
<td>970</td>
<td>.29</td>
<td>1024</td>
</tr>
</tbody>
</table>


b. Determine the Paasche price index for 1999.

14.18 Compute index numbers for the following data using 1960 as the base year.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>PRICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>$2.18</td>
</tr>
<tr>
<td>1965</td>
<td>2.85</td>
</tr>
<tr>
<td>1970</td>
<td>3.47</td>
</tr>
<tr>
<td>1975</td>
<td>4.02</td>
</tr>
<tr>
<td>1980</td>
<td>5.30</td>
</tr>
<tr>
<td>1985</td>
<td>5.11</td>
</tr>
<tr>
<td>1990</td>
<td>5.64</td>
</tr>
<tr>
<td>1995</td>
<td>5.92</td>
</tr>
<tr>
<td>2000</td>
<td>6.08</td>
</tr>
</tbody>
</table>

14.19 Compute unweighted aggregate price index numbers for each of the given years using 1995 as the base year.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.21</td>
<td>3.37</td>
<td>3.80</td>
<td>3.73</td>
<td>3.65</td>
</tr>
<tr>
<td>2</td>
<td>.51</td>
<td>.55</td>
<td>.68</td>
<td>.62</td>
<td>.59</td>
</tr>
<tr>
<td>3</td>
<td>.83</td>
<td>.90</td>
<td>.91</td>
<td>1.02</td>
<td>1.06</td>
</tr>
<tr>
<td>4</td>
<td>1.30</td>
<td>1.32</td>
<td>1.35</td>
<td>1.32</td>
<td>1.30</td>
</tr>
<tr>
<td>5</td>
<td>1.67</td>
<td>1.72</td>
<td>1.90</td>
<td>1.99</td>
<td>1.98</td>
</tr>
<tr>
<td>6</td>
<td>.62</td>
<td>.67</td>
<td>.70</td>
<td>.72</td>
<td>.71</td>
</tr>
</tbody>
</table>

**TESTING YOUR UNDERSTANDING**

14.20 The National Association of Realtors releases data on housing affordability. Listed below are the costs of a median-price existing home in the United States over a 15-year period. Convert these raw data into simple index numbers using 1983 as the base year. Compute simple index numbers on the raw data using 1990 as the base year. Observe the difference in index numbers when different base years are used.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>MEDIAN PRICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>$70,300</td>
</tr>
<tr>
<td>1984</td>
<td>72,400</td>
</tr>
<tr>
<td>1985</td>
<td>75,500</td>
</tr>
<tr>
<td>1986</td>
<td>80,300</td>
</tr>
<tr>
<td>1987</td>
<td>85,600</td>
</tr>
<tr>
<td>1988</td>
<td>90,600</td>
</tr>
<tr>
<td>1990</td>
<td>97,500</td>
</tr>
<tr>
<td>1991</td>
<td>99,700</td>
</tr>
<tr>
<td>1992</td>
<td>103,700</td>
</tr>
<tr>
<td>1993</td>
<td>106,800</td>
</tr>
<tr>
<td>1994</td>
<td>109,900</td>
</tr>
<tr>
<td>1995</td>
<td>113,100</td>
</tr>
<tr>
<td>1996</td>
<td>118,200</td>
</tr>
<tr>
<td>1997</td>
<td>124,100</td>
</tr>
<tr>
<td>1998</td>
<td>134,600</td>
</tr>
</tbody>
</table>

14.21 The U.S. Department of Labor reports the prices of some food commodities. Shown here are the average retail price figures for five different food commodities over 3 years. In addition, quantity estimates are included. Use these data and a base year of 1995 to compute un-

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Margarine (lb)</td>
<td>.83</td>
<td>.81</td>
<td>.83</td>
</tr>
<tr>
<td>Shortening (lb)</td>
<td>.89</td>
<td>.87</td>
<td>.87</td>
</tr>
<tr>
<td>Milk (1/2 gal)</td>
<td>1.43</td>
<td>1.56</td>
<td>1.59</td>
</tr>
<tr>
<td>Cola (2 l)</td>
<td>1.05</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>Potato chips</td>
<td>3.01</td>
<td>3.06</td>
<td>3.13</td>
</tr>
</tbody>
</table>


14.23 The U.S. Department of Agriculture publishes data on livestock in the United States. The following table gives the number of all cattle on U.S. farms for each year from 1900 through 1998 (data are in 1000s). Use 1900 as the base year and calculate simple index numbers for these data. Now use 1960 as the base year and calculate simple index numbers for these data. Compare the results. Is there another year you would select as base year? If so, why?

<table>
<thead>
<tr>
<th>YEAR</th>
<th>NUMBER OF ALL CATTLE (1000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>59,739</td>
</tr>
<tr>
<td>1910</td>
<td>58,993</td>
</tr>
<tr>
<td>1920</td>
<td>70,400</td>
</tr>
<tr>
<td>1930</td>
<td>61,003</td>
</tr>
<tr>
<td>1940</td>
<td>68,309</td>
</tr>
<tr>
<td>1950</td>
<td>77,963</td>
</tr>
<tr>
<td>1960</td>
<td>96,236</td>
</tr>
<tr>
<td>1970</td>
<td>112,369</td>
</tr>
<tr>
<td>1980</td>
<td>111,242</td>
</tr>
<tr>
<td>1985</td>
<td>109,582</td>
</tr>
<tr>
<td>1990</td>
<td>98,162</td>
</tr>
<tr>
<td>1991</td>
<td>98,896</td>
</tr>
<tr>
<td>1992</td>
<td>99,559</td>
</tr>
<tr>
<td>1993</td>
<td>99,176</td>
</tr>
<tr>
<td>1994</td>
<td>100,974</td>
</tr>
<tr>
<td>1995</td>
<td>102,785</td>
</tr>
<tr>
<td>1996</td>
<td>103,548</td>
</tr>
<tr>
<td>1997</td>
<td>101,656</td>
</tr>
<tr>
<td>1998</td>
<td>99,744</td>
</tr>
</tbody>
</table>

14.24 The U.S. Department of Agriculture's Economic Research Service publishes information about the prices of foods in the United States for both retail and farm markets. Shown here are figures released by the Department of Agriculture for four types of fruits and vegetables for 1980, 1985, 1990, and 1995. Using 1980 as the base year, compute unweighted aggregate price indexes for these years for this market basket of foods. Recompute the unweighted aggregate price indexes using 1990 as the base year. Compare the results of the two computations. All prices of foods are computed per pound.
### ANALYZING THE DATABASES

Use the international labor database to answer Questions 1 through 3.

1. Using 1960 as a base year, compute the simple index numbers for the average number of weekly hours worked in the United States. Based on these index numbers, how has the number of hours changed over the years? What trends, if any, do you see in these data?

2. Compute the simple index numbers for the average number of weekly hours worked in Japan from 1960 through 1998, using the following as base years.
   a. 1960
   b. 1975
   c. 1998

3. Calculate the simple index numbers for the average number of weekly hours worked in Germany from 1960 through 1998, using 1960 as a base year. Describe what you find in examining these values. Do you see any trends in the data?

4. Using the agricultural time-series database and month 1 as the base period, determine the index numbers for onions. What do you observe? Is there a continuous increase in index numbers? When do the index numbers peak?
CALCULATING THE STATISTICS

15.21 Following are the average yields of long-term new corporate bonds over a several-month period published by the Office of Market Finance of the U.S. Department of the Treasury.

<table>
<thead>
<tr>
<th>MONTH</th>
<th>YIELD</th>
<th>MONTH</th>
<th>YIELD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.08</td>
<td>13</td>
<td>7.91</td>
</tr>
<tr>
<td>2</td>
<td>10.05</td>
<td>14</td>
<td>7.73</td>
</tr>
<tr>
<td>3</td>
<td>9.24</td>
<td>15</td>
<td>7.39</td>
</tr>
<tr>
<td>4</td>
<td>9.23</td>
<td>16</td>
<td>7.48</td>
</tr>
<tr>
<td>5</td>
<td>9.69</td>
<td>17</td>
<td>7.52</td>
</tr>
<tr>
<td>6</td>
<td>9.55</td>
<td>18</td>
<td>7.48</td>
</tr>
<tr>
<td>7</td>
<td>9.37</td>
<td>19</td>
<td>7.35</td>
</tr>
<tr>
<td>8</td>
<td>8.55</td>
<td>20</td>
<td>7.04</td>
</tr>
<tr>
<td>9</td>
<td>8.36</td>
<td>21</td>
<td>6.88</td>
</tr>
<tr>
<td>10</td>
<td>8.59</td>
<td>22</td>
<td>6.88</td>
</tr>
<tr>
<td>11</td>
<td>7.99</td>
<td>23</td>
<td>7.17</td>
</tr>
<tr>
<td>12</td>
<td>8.12</td>
<td>24</td>
<td>7.22</td>
</tr>
</tbody>
</table>

15.23 The U.S. Department of Commerce publishes a series of census documents referred to as Current Industrial Reports. Included in these documents are the Manufacturers' Shipment, Inventories, and Orders: 1991–1995. Displayed here is a portion of these data representing the shipments of chemicals and allied products from January 1991 through December 1995. Use time-series decomposition methods to develop the seasonal indexes for these data.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>QUANTITY</th>
<th>YEAR</th>
<th>QUANTITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>3,654</td>
<td>1989</td>
<td>6,204</td>
</tr>
<tr>
<td>1981</td>
<td>3,547</td>
<td>1990</td>
<td>7,041</td>
</tr>
<tr>
<td>1982</td>
<td>3,285</td>
<td>1991</td>
<td>7,031</td>
</tr>
<tr>
<td>1983</td>
<td>3,238</td>
<td>1992</td>
<td>7,618</td>
</tr>
<tr>
<td>1984</td>
<td>3,320</td>
<td>1993</td>
<td>8,214</td>
</tr>
<tr>
<td>1985</td>
<td>3,294</td>
<td>1994</td>
<td>7,936</td>
</tr>
<tr>
<td>1986</td>
<td>3,393</td>
<td>1995</td>
<td>7,667</td>
</tr>
<tr>
<td>1987</td>
<td>3,946</td>
<td>1996</td>
<td>7,474</td>
</tr>
<tr>
<td>1988</td>
<td>4,588</td>
<td>1997</td>
<td>7,248</td>
</tr>
</tbody>
</table>

a. Explore trends in these data by using regression trend analysis. How strong are the models? Is the quadratic model significantly stronger than the linear trend model?
b. Use a 4-month moving average to forecast values for each of the ensuing months.
c. Use simple exponential smoothing to forecast values for each of the ensuing months. Let $\alpha = .3$ and then let $\alpha = .7$. Which weight produces better forecasts?
d. Compute MAD for the forecasts obtained in (b) and (c) and compare the results.
e. Perform decomposition on these data. Let the seasonal effects have four periods and use a linear trend.

TESTING YOUR UNDERSTANDING

15.22 Following are data on the quantity (million pounds) of the U.S. domestic fishing catch for human food from 1980 through 1997. The data are published by the U.S. National Oceanic and Atmosphere Administration.

<table>
<thead>
<tr>
<th>TIME PERIOD</th>
<th>QUANTITY ($BILLION)</th>
</tr>
</thead>
<tbody>
<tr>
<td>January (1991)</td>
<td>23.701</td>
</tr>
<tr>
<td>February</td>
<td>24.189</td>
</tr>
<tr>
<td>March</td>
<td>24.200</td>
</tr>
<tr>
<td>April</td>
<td>24.971</td>
</tr>
<tr>
<td>May</td>
<td>24.560</td>
</tr>
<tr>
<td>June</td>
<td>24.992</td>
</tr>
<tr>
<td>July</td>
<td>22.566</td>
</tr>
<tr>
<td>August</td>
<td>24.037</td>
</tr>
<tr>
<td>September</td>
<td>25.047</td>
</tr>
<tr>
<td>October</td>
<td>24.115</td>
</tr>
<tr>
<td>November</td>
<td>23.034</td>
</tr>
<tr>
<td>December</td>
<td>22.590</td>
</tr>
<tr>
<td>January (1992)</td>
<td>23.347</td>
</tr>
<tr>
<td>February</td>
<td>24.122</td>
</tr>
<tr>
<td>March</td>
<td>25.282</td>
</tr>
<tr>
<td>April</td>
<td>25.426</td>
</tr>
<tr>
<td>May</td>
<td>25.185</td>
</tr>
<tr>
<td>June</td>
<td>26.486</td>
</tr>
<tr>
<td>July</td>
<td>24.088</td>
</tr>
<tr>
<td>August</td>
<td>24.672</td>
</tr>
<tr>
<td>September</td>
<td>26.072</td>
</tr>
<tr>
<td>October</td>
<td>24.328</td>
</tr>
<tr>
<td>November</td>
<td>23.826</td>
</tr>
<tr>
<td>December</td>
<td>24.373</td>
</tr>
<tr>
<td>January (1993)</td>
<td>24.207</td>
</tr>
<tr>
<td>February</td>
<td>25.772</td>
</tr>
<tr>
<td>March</td>
<td>27.591</td>
</tr>
<tr>
<td>April</td>
<td>26.958</td>
</tr>
<tr>
<td>May</td>
<td>25.920</td>
</tr>
<tr>
<td>June</td>
<td>28.460</td>
</tr>
</tbody>
</table>

a. Use a 3-year moving average to forecast the quantity of fish for the years 1983 through 1997 for these data. Compute the error of each forecast and then determine the mean absolute deviation of error for the forecast.
b. Use exponential smoothing and $\alpha = .2$ to forecast the data from 1983 through 1997. Let the forecast for 1981 equal the actual value for 1980. Compute the error of each forecast and then determine the mean absolute deviation of error for the forecast.
c. Compare the results obtained in (a) and (b) by using MAD. Which technique seems to perform better? Why?
15.24 Use the seasonal indexes computed in Problem 15.23 to deseasonalize the data from Problem 15.23.

15.25 Determine the trend for the data in Problem 15.23 using the deseasonalized data from Problem 15.24. Explore both a linear and a quadratic model in an attempt to develop the better trend model.

15.26 Use the linear trend results obtained in Problem 15.25 and 12-month centered moving averages in Problem 15.23 to isolate the cyclical effects of the data in Problem 15.23.

15.27 The National Cable Television Association publishes data on the cable television market. Shown here are the number of basic cable subscribers and as percentage of household with TVs from the year 1976 to 1997. Develop a regression model to predict the number of basic cable subscribers from the variable as percentage of households with TVs using these data. Use this model to predict the number of basic cable subscribers if the value of the variable as percentage of households with TVs is 55%. Discuss the strength of the regression model.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>BASIC CABLE SUBSCRIBERS</th>
<th>AS PERCENTAGE OF HOUSEHOLDS WITH TVS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>10,787,970</td>
<td>15.1</td>
</tr>
<tr>
<td>1977</td>
<td>12,168,450</td>
<td>16.6</td>
</tr>
<tr>
<td>1978</td>
<td>13,391,910</td>
<td>17.9</td>
</tr>
<tr>
<td>1979</td>
<td>14,814,380</td>
<td>19.4</td>
</tr>
<tr>
<td>1980</td>
<td>17,671,490</td>
<td>22.6</td>
</tr>
<tr>
<td>1981</td>
<td>23,219,200</td>
<td>28.3</td>
</tr>
<tr>
<td>1982</td>
<td>29,340,570</td>
<td>35.0</td>
</tr>
<tr>
<td>1983</td>
<td>34,113,790</td>
<td>40.5</td>
</tr>
<tr>
<td>1984</td>
<td>37,290,870</td>
<td>43.7</td>
</tr>
<tr>
<td>1985</td>
<td>39,872,520</td>
<td>46.2</td>
</tr>
<tr>
<td>1986</td>
<td>42,237,140</td>
<td>48.1</td>
</tr>
<tr>
<td>1987</td>
<td>44,970,880</td>
<td>50.5</td>
</tr>
<tr>
<td>1988</td>
<td>48,636,520</td>
<td>53.8</td>
</tr>
<tr>
<td>1989</td>
<td>52,564,470</td>
<td>57.1</td>
</tr>
<tr>
<td>1990</td>
<td>54,871,330</td>
<td>59.0</td>
</tr>
<tr>
<td>1991</td>
<td>55,786,390</td>
<td>60.6</td>
</tr>
<tr>
<td>1992</td>
<td>57,211,660</td>
<td>61.5</td>
</tr>
<tr>
<td>1993</td>
<td>58,834,440</td>
<td>62.5</td>
</tr>
<tr>
<td>1994</td>
<td>60,483,600</td>
<td>63.4</td>
</tr>
<tr>
<td>1995</td>
<td>62,956,470</td>
<td>65.7</td>
</tr>
<tr>
<td>1996</td>
<td>64,654,180</td>
<td>66.7</td>
</tr>
<tr>
<td>1997</td>
<td>65,929,420</td>
<td>67.3</td>
</tr>
</tbody>
</table>

15.29 Use the data and the regression model developed in Problem 15.27 to compute a Durbin-Watson test to determine whether significant autocorrelation is present. Let $\alpha = .05$.

15.30 In the Survey of Current Business, the U.S. Department of Commerce publishes data on farm commodity prices. Given are the cotton prices from November of year 1 through February of year 4. The prices are indexes with a base of 100 from the period of 1910 through 1914. Use these data to develop autoregression models for a 1-month lag and a 4-month lag. Compare the results of these two models. Which model seems to yield better predictions?

<table>
<thead>
<tr>
<th>TIME PERIOD</th>
<th>COTTON PRICES</th>
</tr>
</thead>
<tbody>
<tr>
<td>November (year 1)</td>
<td>552</td>
</tr>
<tr>
<td>December</td>
<td>519</td>
</tr>
<tr>
<td>January (year 2)</td>
<td>505</td>
</tr>
<tr>
<td>February</td>
<td>512</td>
</tr>
<tr>
<td>March</td>
<td>541</td>
</tr>
<tr>
<td>April</td>
<td>549</td>
</tr>
<tr>
<td>May</td>
<td>552</td>
</tr>
<tr>
<td>June</td>
<td>526</td>
</tr>
<tr>
<td>July</td>
<td>531</td>
</tr>
<tr>
<td>August</td>
<td>545</td>
</tr>
<tr>
<td>September</td>
<td>549</td>
</tr>
<tr>
<td>October</td>
<td>570</td>
</tr>
<tr>
<td>November</td>
<td>576</td>
</tr>
<tr>
<td>December</td>
<td>568</td>
</tr>
<tr>
<td>January (year 4)</td>
<td>436</td>
</tr>
<tr>
<td>February</td>
<td>419</td>
</tr>
</tbody>
</table>

15.31 The Board of Governors of the Federal Reserve System publishes data on mortgage debt outstanding by type of property and holder. Below are the amounts of residential nonfarm debt (in billions of dollars) held by savings institutions in the United States over a 10-year period.
Use these data to develop an autoregression model with a 1-period lag. Discuss the strength of the model.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>DEBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>529</td>
</tr>
<tr>
<td>2</td>
<td>554</td>
</tr>
<tr>
<td>3</td>
<td>559</td>
</tr>
<tr>
<td>4</td>
<td>602</td>
</tr>
<tr>
<td>5</td>
<td>672</td>
</tr>
<tr>
<td>6</td>
<td>669</td>
</tr>
<tr>
<td>7</td>
<td>600</td>
</tr>
<tr>
<td>8</td>
<td>538</td>
</tr>
<tr>
<td>9</td>
<td>490</td>
</tr>
<tr>
<td>10</td>
<td>470</td>
</tr>
</tbody>
</table>

15.32 The U.S. Department of Commerce publishes data on industrial machinery and equipment. Shown here are the shipments (in billions of dollars) of industrial machinery and equipment from the first quarter of year 1 through the fourth quarter of year 6. Use these data to determine the seasonal indexes for the data through time-series decomposition methods. Use the four-quarter centered moving average in the computations.

<table>
<thead>
<tr>
<th>TIME PERIOD</th>
<th>INDUSTRIAL MACHINERY AND EQUIPMENT SHIPMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quarter (year 1)</td>
<td>54.019</td>
</tr>
<tr>
<td>2nd quarter</td>
<td>56.495</td>
</tr>
<tr>
<td>3rd quarter</td>
<td>50.169</td>
</tr>
<tr>
<td>4th quarter</td>
<td>52.891</td>
</tr>
<tr>
<td>1st quarter (year 2)</td>
<td>51.915</td>
</tr>
<tr>
<td>2nd quarter</td>
<td>55.101</td>
</tr>
<tr>
<td>3rd quarter</td>
<td>53.419</td>
</tr>
<tr>
<td>4th quarter</td>
<td>57.236</td>
</tr>
<tr>
<td>1st quarter (year 3)</td>
<td>57.063</td>
</tr>
<tr>
<td>2nd quarter</td>
<td>62.488</td>
</tr>
<tr>
<td>3rd quarter</td>
<td>60.373</td>
</tr>
<tr>
<td>4th quarter</td>
<td>63.334</td>
</tr>
<tr>
<td>1st quarter (year 4)</td>
<td>62.723</td>
</tr>
<tr>
<td>2nd quarter</td>
<td>68.380</td>
</tr>
<tr>
<td>3rd quarter</td>
<td>63.256</td>
</tr>
<tr>
<td>4th quarter</td>
<td>66.446</td>
</tr>
<tr>
<td>1st quarter (year 5)</td>
<td>65.445</td>
</tr>
<tr>
<td>2nd quarter</td>
<td>68.011</td>
</tr>
<tr>
<td>3rd quarter</td>
<td>63.245</td>
</tr>
<tr>
<td>4th quarter</td>
<td>66.872</td>
</tr>
<tr>
<td>1st quarter (year 6)</td>
<td>59.714</td>
</tr>
<tr>
<td>2nd quarter</td>
<td>63.590</td>
</tr>
<tr>
<td>3rd quarter</td>
<td>58.088</td>
</tr>
<tr>
<td>4th quarter</td>
<td>61.443</td>
</tr>
</tbody>
</table>

15.33 Use the seasonal indexes computed in Problem 15.32 to deseasonalize the data in Problem 15.32.

15.34 Use both a linear and quadratic model to explore trends in the deseasonalized data from Problem 15.33. Which model seems to produce a better fit of the data?

15.35 Use the quadratic trend model developed in Problem 15.34 and the 4-month centered moving averages developed in Problem 15.32 to determine the cyclical effects in the data from Problem 15.32.

15.36 The data shown here, from the Investment Company Institute, show that the equity fund assets of mutual funds have been growing since 1981. At the same time, the assets of mutual funds in taxable money markets have been increasing since 1980. Use these data to develop a regression model to forecast the equity fund assets by the taxable money market assets. All figures are given in billion-dollar units. Conduct a Durbin-Watson test on the data and the regression model to determine whether significant autocorrelation is present. Let $\alpha = .01$.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>EQUITY FUNDS</th>
<th>TAXABLE MONEY MARKETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>44.4</td>
<td>74.5</td>
</tr>
<tr>
<td>1981</td>
<td>41.2</td>
<td>181.9</td>
</tr>
<tr>
<td>1982</td>
<td>53.7</td>
<td>206.6</td>
</tr>
<tr>
<td>1983</td>
<td>77.0</td>
<td>162.5</td>
</tr>
<tr>
<td>1984</td>
<td>83.1</td>
<td>209.7</td>
</tr>
<tr>
<td>1985</td>
<td>116.9</td>
<td>207.5</td>
</tr>
<tr>
<td>1986</td>
<td>161.5</td>
<td>228.3</td>
</tr>
<tr>
<td>1987</td>
<td>180.7</td>
<td>254.7</td>
</tr>
<tr>
<td>1988</td>
<td>194.8</td>
<td>272.3</td>
</tr>
<tr>
<td>1989</td>
<td>249.0</td>
<td>358.7</td>
</tr>
<tr>
<td>1990</td>
<td>245.8</td>
<td>414.7</td>
</tr>
<tr>
<td>1991</td>
<td>411.6</td>
<td>452.6</td>
</tr>
<tr>
<td>1992</td>
<td>522.8</td>
<td>451.4</td>
</tr>
<tr>
<td>1993</td>
<td>749.0</td>
<td>461.9</td>
</tr>
<tr>
<td>1994</td>
<td>866.4</td>
<td>500.4</td>
</tr>
<tr>
<td>1995</td>
<td>1269.0</td>
<td>629.7</td>
</tr>
<tr>
<td>1996</td>
<td>1750.9</td>
<td>761.8</td>
</tr>
<tr>
<td>1997</td>
<td>2399.3</td>
<td>898.1</td>
</tr>
</tbody>
</table>

15.37 The purchasing-power value figures for the minimum wage in 1997 dollars for the years 1980 through 1997 are shown here. Use these data and exponential smoothing to develop forecasts for the years 1981 through 1997. Try $\alpha = .1, .5,$ and .8, and compare the results using MAPE. Discuss your findings. Select the value of alpha that worked best and use your exponential smoothing results to predict the figure for 1998.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>PURCHASING POWER</th>
<th>PURCHASING POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>$6.04$</td>
<td>1989</td>
</tr>
<tr>
<td>1981</td>
<td>5.92</td>
<td>1990</td>
</tr>
<tr>
<td>1982</td>
<td>5.57</td>
<td>1991</td>
</tr>
<tr>
<td>1983</td>
<td>5.40</td>
<td>1992</td>
</tr>
<tr>
<td>1984</td>
<td>5.17</td>
<td>1993</td>
</tr>
<tr>
<td>1985</td>
<td>5.00</td>
<td>1994</td>
</tr>
<tr>
<td>1986</td>
<td>4.91</td>
<td>1995</td>
</tr>
<tr>
<td>1987</td>
<td>4.73</td>
<td>1996</td>
</tr>
<tr>
<td>1988</td>
<td>4.55</td>
<td>1997</td>
</tr>
</tbody>
</table>

**INTERPRETING THE OUTPUT**

15.38 Shown here is the Excel output for a regression analysis to predict the number of business bankruptcy filings
over a 16-year period by the number of consumer bankruptcy filings. How strong is the model? Note the residuals. Compute a Durbin-Watson statistic from the data and discuss the presence of autocorrelation in this model.

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>364069877.4</td>
<td>3.64E+08</td>
<td>5.44</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>14</td>
<td>936737379.6</td>
<td>66909813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>1300807257</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Standard Coefficients**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>75532.436</td>
<td>4980.088</td>
<td>15.17</td>
</tr>
<tr>
<td>Consumer Bankruptcies</td>
<td>-0.016</td>
<td>0.007</td>
<td>-2.33</td>
</tr>
</tbody>
</table>

**Observation**

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted Business Bankruptcies</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70638.58</td>
<td>-1338.58</td>
</tr>
<tr>
<td>2</td>
<td>71024.28</td>
<td>-8588.28</td>
</tr>
<tr>
<td>3</td>
<td>71054.61</td>
<td>-7050.61</td>
</tr>
<tr>
<td>4</td>
<td>70161.99</td>
<td>1115.01</td>
</tr>
<tr>
<td>5</td>
<td>68462.72</td>
<td>12772.28</td>
</tr>
<tr>
<td>6</td>
<td>67733.25</td>
<td>14712.75</td>
</tr>
<tr>
<td>7</td>
<td>66882.45</td>
<td>-3029.45</td>
</tr>
<tr>
<td>8</td>
<td>65834.05</td>
<td>-2599.05</td>
</tr>
<tr>
<td>9</td>
<td>64230.61</td>
<td>622.39</td>
</tr>
<tr>
<td>10</td>
<td>61801.70</td>
<td>9747.30</td>
</tr>
<tr>
<td>11</td>
<td>61354.16</td>
<td>9288.84</td>
</tr>
<tr>
<td>12</td>
<td>62738.76</td>
<td>-434.76</td>
</tr>
<tr>
<td>13</td>
<td>63249.36</td>
<td>-10875.36</td>
</tr>
<tr>
<td>14</td>
<td>61767.01</td>
<td>-9808.01</td>
</tr>
<tr>
<td>15</td>
<td>57826.69</td>
<td>-4277.69</td>
</tr>
<tr>
<td>16</td>
<td>54283.80</td>
<td>-256.80</td>
</tr>
</tbody>
</table>

15.39 A company's production data for 40 quarters (10 years) is analyzed using MINITAB's forecasting graphical analysis. The output is shown here. Study the output and discuss the forecasts, the size of errors, and the effects that were analyzed.
Seasonal Analysis for Production

Seasonal Indexes

Original Data, by Seasonal Period

Percent Variation, by Seasonal Period

Residuals, by Seasonal Period

Component Analysis for Production

Original Data

Detrended Data

Seasonally Adjusted Data

Seasonally Adjusted and Detrended Data
CALCULATING THE STATISTICS

16.56 Use a chi-square goodness-of-fit test to determine whether the following observed frequencies are distributed the same as the expected frequencies. Let $\alpha = .01$.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>$f_0$</th>
<th>$f_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>214</td>
<td>206</td>
</tr>
<tr>
<td>2</td>
<td>235</td>
<td>232</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>268</td>
</tr>
<tr>
<td>4</td>
<td>281</td>
<td>284</td>
</tr>
<tr>
<td>5</td>
<td>264</td>
<td>268</td>
</tr>
<tr>
<td>6</td>
<td>254</td>
<td>232</td>
</tr>
<tr>
<td>7</td>
<td>211</td>
<td>206</td>
</tr>
</tbody>
</table>

16.57 Use the chi-square contingency analysis to test to determine whether variable 1 is independent of variable 2. Use a 5% level of significance.

<table>
<thead>
<tr>
<th>Variable 2</th>
<th>12</th>
<th>23</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable 1</td>
<td>8</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

16.58 Use the runs test to determine whether the sample is random or not. Let $\alpha = .05$.

1 1 1 1 1 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 1 2 1 1 1 1 2 2 2

16.59 Use the Mann-Whitney $U$ test and $\alpha = .01$ to determine whether there is a significant difference between the populations represented by the two samples given here.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>573</td>
<td>547</td>
</tr>
<tr>
<td>532</td>
<td>566</td>
</tr>
<tr>
<td>544</td>
<td>551</td>
</tr>
<tr>
<td>565</td>
<td>538</td>
</tr>
<tr>
<td>540</td>
<td>557</td>
</tr>
<tr>
<td>548</td>
<td>560</td>
</tr>
<tr>
<td>556</td>
<td>557</td>
</tr>
<tr>
<td>523</td>
<td>547</td>
</tr>
</tbody>
</table>

16.60 Use the Wilcoxon matched-pairs signed rank test to determine whether there is a significant difference between the related populations represented by the matched pairs given here. Assume $\alpha = .05$.

<table>
<thead>
<tr>
<th>GROUP 1</th>
<th>GROUP 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6</td>
<td>6.4</td>
</tr>
<tr>
<td>1.3</td>
<td>1.5</td>
</tr>
<tr>
<td>4.7</td>
<td>4.6</td>
</tr>
<tr>
<td>3.8</td>
<td>4.3</td>
</tr>
<tr>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>5.5</td>
<td>6.0</td>
</tr>
<tr>
<td>5.1</td>
<td>5.2</td>
</tr>
<tr>
<td>4.6</td>
<td>4.5</td>
</tr>
<tr>
<td>3.7</td>
<td>4.5</td>
</tr>
</tbody>
</table>

16.61 Use the Kruskal-Wallis test and $\alpha = .01$ to determine whether the four groups come from different populations.

<table>
<thead>
<tr>
<th>GROUP 1</th>
<th>GROUP 2</th>
<th>GROUP 3</th>
<th>GROUP 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

16.62 Use the Friedman test to determine whether the treatment groups come from different populations. Let $\alpha = .05$.

<table>
<thead>
<tr>
<th>BLOCK</th>
<th>GROUP 1</th>
<th>GROUP 2</th>
<th>GROUP 3</th>
<th>GROUP 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>17</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>26</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>11</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>16</td>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

16.63 Compute a Spearman's rank correlation to determine the degree of association between the two variables.

<table>
<thead>
<tr>
<th>VARIABLE 1</th>
<th>VARIABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>87</td>
</tr>
<tr>
<td>129</td>
<td>89</td>
</tr>
<tr>
<td>133</td>
<td>84</td>
</tr>
<tr>
<td>147</td>
<td>79</td>
</tr>
<tr>
<td>156</td>
<td>70</td>
</tr>
<tr>
<td>179</td>
<td>64</td>
</tr>
<tr>
<td>183</td>
<td>67</td>
</tr>
<tr>
<td>190</td>
<td>71</td>
</tr>
</tbody>
</table>

TESTING YOUR UNDERSTANDING

16.64 Is a manufacturer's geographic location independent of type of customer? Use the following data for companies with primarily industrial customers and companies with primarily retail customers to test this question. Let $\alpha = .10$.

<table>
<thead>
<tr>
<th>Geographic Location</th>
<th>Northeast</th>
<th>West</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customer Type</strong></td>
<td><strong>Industrial Customer</strong></td>
<td>230</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td><strong>Retail Customer</strong></td>
<td>185</td>
<td>143</td>
</tr>
</tbody>
</table>
16.65 Commercial fish raising is a growing industry in the United States. What makes fish raised commercially grow faster and larger? Suppose that a fish industry study is conducted over the three summer months in an effort to determine whether the amount of water allotted per fish makes any difference in the speed with which the fish grow. The following data represent the inches of growth of marked catfish in fish farms for different volumes of water per fish. Use $\alpha = .01$ to test whether there is a significant difference in fish growth by volume of allotted water.

<table>
<thead>
<tr>
<th>KIND OF COOKIE</th>
<th>OBSERVED FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chocolate chip</td>
<td>189</td>
</tr>
<tr>
<td>Peanut butter</td>
<td>168</td>
</tr>
<tr>
<td>Cheese cracker</td>
<td>155</td>
</tr>
<tr>
<td>Lemon flavored</td>
<td>161</td>
</tr>
<tr>
<td>Chocolate mint</td>
<td>216</td>
</tr>
<tr>
<td>Vanilla filled</td>
<td>165</td>
</tr>
</tbody>
</table>

16.66 Because of their success, some television commercials have lives of several years. When a commercial is successful for several years, how much correlation is there between the ranking of the commercial one year and its ranking the next? Listed are the top six companies with successful television commercials in 1997 according to Video Storyboard Tests, Inc., along with rankings in 1997 and 1996 relative rankings. Determine the correlation of the rankings for these two years.

<table>
<thead>
<tr>
<th>BRAND</th>
<th>1997 RANKING</th>
<th>1996 RELATIVE RANKING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nissan</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Budweiser</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Pepsi</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Milk</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>McDonald's</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Coca-Cola</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

16.67 Manchester Partners International claims that 60% of the banking executives who lose their job stay in banking whereas 40% leave banking. Suppose 40 people who have lost their job as a banking executive are contacted and are asked whether they are still in banking or not. The results follow. Test to determine whether this sample appears to be random on the basis of the sequence of those who have left banking and those who have not. Let L denote left banking and S denote stay. Let $\alpha = .05$.

| S S L S L S L S S S S L S S L L L S L S S S S L L S S L S S L S |

16.68 A national youth organization sells six different kinds of cookies during its annual cookie campaign. A local leader is curious about whether national sales of the six kinds of cookies are uniformly distributed. He randomly selects the amounts of each kind of cookies sold from five youths and combines them into the observed data shown. Use $\alpha = .05$ to determine whether the data indicate that sales for these six kinds of cookies are uniformly distributed.

16.69 Three machines produce the same part. Ten different machine operators work these machines. A quality team wants to determine whether the machines are producing parts that are significantly different from each other in weight. The team devises an experimental design in which a random part is selected from each of the ten machine operators on each machine. The results follow. Using alpha of .05, test to determine whether there is a difference in machines.

<table>
<thead>
<tr>
<th>OPERATOR</th>
<th>MACHINE 1</th>
<th>MACHINE 2</th>
<th>MACHINE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>231</td>
<td>229</td>
<td>234</td>
</tr>
<tr>
<td>2</td>
<td>233</td>
<td>232</td>
<td>231</td>
</tr>
<tr>
<td>3</td>
<td>229</td>
<td>233</td>
<td>230</td>
</tr>
<tr>
<td>4</td>
<td>232</td>
<td>235</td>
<td>231</td>
</tr>
<tr>
<td>5</td>
<td>235</td>
<td>228</td>
<td>232</td>
</tr>
<tr>
<td>6</td>
<td>234</td>
<td>237</td>
<td>231</td>
</tr>
<tr>
<td>7</td>
<td>236</td>
<td>233</td>
<td>230</td>
</tr>
<tr>
<td>8</td>
<td>230</td>
<td>229</td>
<td>227</td>
</tr>
<tr>
<td>9</td>
<td>228</td>
<td>230</td>
<td>229</td>
</tr>
<tr>
<td>10</td>
<td>237</td>
<td>238</td>
<td>234</td>
</tr>
</tbody>
</table>

16.70 A researcher interviewed 2067 people and asked whether they were the primary decision makers in the household when buying a new car last year. Two hundred seven were men and had bought a new car last year. Sixty-five were women and had bought a new car last year. Eight hundred eleven of the responses were from men who did not buy a car last year. Nine hundred eighty-four were from women who did not buy a car last year. Use these data to determine whether gender is independent of being a major decision maker in purchasing a car last year. Let $\alpha = .05$.

16.71 Are random arrivals at a shoe store at the local mall Poisson distributed? Suppose a mall employee researches this question by gathering data for arrivals during one-minute intervals on a weekday between 6:30 PM. and 8:00 PM. The data obtained follow. Use $\alpha = .05$ to determine whether the observed data seem to be from a Poisson distribution.
ARRIVALS PER MINUTE  |  OBSERVED FREQUENCY
---------------------|---------------------
0                    | 26                  
1                    | 40                  
2                    | 57                  
3                    | 32                  
4                    | 17                  
5                    | 12                  
6                    | 8                   

16.72 In some fire-fighting organizations, you must serve as a fire fighter for some period of time before you can become part of the emergency medical service arm of the organization. Does that mean EMS workers are older, on average, than traditional fire fighters? Use the data shown and $\alpha = .05$ to test whether EMS workers are significantly older than fire fighters. Assume the two groups are independent and you do not want to use a $t$ test to analyze the data.

<table>
<thead>
<tr>
<th>FIRE FIGHTERS</th>
<th>EMS WORKERS</th>
<th>FIRE FIGHTERS</th>
<th>EMS WORKERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>27</td>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>37</td>
<td>29</td>
<td>24</td>
<td>33</td>
</tr>
<tr>
<td>28</td>
<td>30</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>25</td>
<td>33</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>41</td>
<td>28</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>36</td>
<td>36</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

16.73 Automobile dealers usually advertise in the yellow pages of the telephone book. Sometimes they have to pay to be listed in the white pages, and some dealerships opt to save money by omitting that listing, assuming most people will use the yellow pages to find the telephone number. A 2-year study is conducted with 20 car dealerships where in one year the dealer is listed in the white pages and the other year it is not. Ten of the dealerships are listed in the white pages the first year and the other 10 are listed there in the second year in an attempt to control for economic cycles. The following data represent the numbers of units sold per year. Is there a significant difference between the number of units sold when the dealership is listed in the white pages and the number sold when it is not listed? Assume all companies are continuously listed in the yellow pages, that the $t$ test is not appropriate, and that $\alpha = .01$.

<table>
<thead>
<tr>
<th>DEALER</th>
<th>WITH LISTING</th>
<th>WITHOUT LISTING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1180</td>
<td>1209</td>
</tr>
<tr>
<td>2</td>
<td>874</td>
<td>902</td>
</tr>
<tr>
<td>3</td>
<td>1071</td>
<td>862</td>
</tr>
<tr>
<td>4</td>
<td>668</td>
<td>503</td>
</tr>
<tr>
<td>5</td>
<td>889</td>
<td>974</td>
</tr>
<tr>
<td>6</td>
<td>724</td>
<td>675</td>
</tr>
<tr>
<td>7</td>
<td>880</td>
<td>821</td>
</tr>
<tr>
<td>8</td>
<td>482</td>
<td>567</td>
</tr>
<tr>
<td>9</td>
<td>796</td>
<td>602</td>
</tr>
<tr>
<td>10</td>
<td>1207</td>
<td>1097</td>
</tr>
</tbody>
</table>

16.74 According to *Beverage Digest/Maxwell Report*, the distribution of market share for the top six soft drinks in the United States was Coca-Cola Classic 20.6%, Pepsi 14.5%, Diet Coke 8.5%, Mountain Dew 6.3%, Sprite 6.2%, Dr. Pepper 5.9%, and others 38%. Suppose a marketing analyst wants to determine whether this distribution fits that of her geographic region. She randomly surveys 1726 local people and asks them to name their favorite soft drink. The responses are: Classic Coke 361, Pepsi 272, Diet Coke 192, Mountain Dew 121, Sprite 102, Dr. Pepper 94, and others 584. She then tests to determine whether the local distribution of soft-drink preferences is the same or different from the national figures, using $\alpha = .05$. What does she find?

16.75 Suppose you want to take a random sample of GMAT test scores to determine whether there is any significant difference between the GMAT scores for the test given in March and the scores for the test given in June. You gather the following data from a sample of persons who took each test. Use the Mann-Whitney $U$ test to determine whether there is a significant difference in the two test results. Let $\alpha = .10$.

<table>
<thead>
<tr>
<th>MARCH</th>
<th>JUNE</th>
</tr>
</thead>
<tbody>
<tr>
<td>490</td>
<td>300</td>
</tr>
<tr>
<td>520</td>
<td>420</td>
</tr>
<tr>
<td>550</td>
<td>580</td>
</tr>
<tr>
<td>380</td>
<td>540</td>
</tr>
<tr>
<td>450</td>
<td>560</td>
</tr>
<tr>
<td>460</td>
<td>470</td>
</tr>
<tr>
<td>480</td>
<td>410</td>
</tr>
<tr>
<td>510</td>
<td>500</td>
</tr>
<tr>
<td>500</td>
<td>480</td>
</tr>
<tr>
<td>440</td>
<td>520</td>
</tr>
</tbody>
</table>

16.76 Does impulse buying really increase sales? A market researcher is curious to find out whether the location of packages of chewing gum in a grocery store really has anything to do with volume of gum sales. As a test, gum is moved to a different location in the store every Monday for four weeks (four locations). To control the experiment for type of gum, six different brands are moved around. Sales representatives keep track of how many packs of each type of gum are sold every Monday for the four weeks. The results follow. Test to determine whether
16.77 Does deodorant sell better in a box or without additional packaging? An experiment in a large store is designed in which, for one month, all deodorants are sold packaged in a box and, during a second month, all deodorants are removed from the box and sold without packaging. Is there a significant difference in the number of units of deodorant sold with and without the additional packaging? Let \( \alpha = .05 \).

<table>
<thead>
<tr>
<th>DEODORANT</th>
<th>BOX</th>
<th>NO BOX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>185</td>
<td>170</td>
</tr>
<tr>
<td>2</td>
<td>109</td>
<td>112</td>
</tr>
<tr>
<td>3</td>
<td>92</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>105</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>58</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>161</td>
<td>165</td>
</tr>
<tr>
<td>10</td>
<td>108</td>
<td>82</td>
</tr>
<tr>
<td>11</td>
<td>89</td>
<td>94</td>
</tr>
<tr>
<td>12</td>
<td>123</td>
<td>139</td>
</tr>
<tr>
<td>13</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>14</td>
<td>68</td>
<td>55</td>
</tr>
<tr>
<td>15</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>16</td>
<td>78</td>
<td>52</td>
</tr>
</tbody>
</table>

16.78 Some people drink coffee to relieve stress on the job. Is there a correlation between the number of cups of coffee consumed on the job and perceived job stress? Suppose the data shown represent the number of cups of coffee consumed per week and a stress rating for the job on a scale of 0 to 100 for nine managers in the same industry. Determine the correlation between these two variables, assuming you do not want to use the Pearson product-moment correlation coefficient.

<table>
<thead>
<tr>
<th>CUPS OF COFFEE PER WEEK</th>
<th>JOB STRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>80</td>
</tr>
<tr>
<td>41</td>
<td>85</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
</tr>
<tr>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>34</td>
<td>65</td>
</tr>
<tr>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>

16.79 Are the types of professional jobs held in the computing industry independent of the number of years a person has worked in the industry? Suppose 246 workers are interviewed. Use the results obtained to determine whether type of professional job held in the computer industry is independent of years worked in the industry. Let \( \alpha = .01 \).

<table>
<thead>
<tr>
<th>Professional Position</th>
<th>Manager</th>
<th>Programmer</th>
<th>Operator</th>
<th>Systems Analyst</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3</td>
<td>6</td>
<td>37</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>4-8</td>
<td>28</td>
<td>16</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>More than 8</td>
<td>47</td>
<td>10</td>
<td>12</td>
<td>19</td>
</tr>
</tbody>
</table>

16.80 A Gallup/Air Transport Association survey showed that in a recent year, 52% of all air trips were for pleasure/personal and 48% were for business. Suppose the organization randomly samples 30 air travelers and asks them to state the purpose of their trip. The results are shown here with B denoting business and P denoting personal. Test the sequence of these data to determine whether the data are random. Let \( \alpha = .05 \).


16.81 Does a statistics course improve a student's mathematics skills, as measured by a national test? Suppose a random sample of 13 students takes the same national mathematics examination just prior to enrolling in a statistics course and just after completing the course. Listed are the students' quantitative scores from both examinations. Use \( \alpha = .01 \) to determine whether the scores after the statistics course are significantly higher than the scores before.

<table>
<thead>
<tr>
<th>STUDENT</th>
<th>BEFORE</th>
<th>AFTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>430</td>
<td>465</td>
</tr>
<tr>
<td>2</td>
<td>485</td>
<td>475</td>
</tr>
<tr>
<td>3</td>
<td>520</td>
<td>535</td>
</tr>
<tr>
<td>4</td>
<td>360</td>
<td>410</td>
</tr>
<tr>
<td>5</td>
<td>440</td>
<td>425</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>505</td>
</tr>
<tr>
<td>7</td>
<td>425</td>
<td>450</td>
</tr>
<tr>
<td>8</td>
<td>470</td>
<td>480</td>
</tr>
<tr>
<td>9</td>
<td>515</td>
<td>520</td>
</tr>
<tr>
<td>10</td>
<td>430</td>
<td>430</td>
</tr>
<tr>
<td>11</td>
<td>450</td>
<td>460</td>
</tr>
<tr>
<td>12</td>
<td>495</td>
<td>500</td>
</tr>
<tr>
<td>13</td>
<td>540</td>
<td>530</td>
</tr>
</tbody>
</table>

16.82 A study by Market Facts/TeleNation for Personnel Decisions International (PDI) found that the average workweek is getting longer for U.S. full-time workers. Forty-three percent of the responding workers in the survey cited "more work, more business" as the number one reason for
this increase in workweek. Suppose you want to test this figure in California to determine whether California workers feel the same way. A random sample of 315 California full-time workers whose workweek has been getting longer is chosen. They are offered a selection of possible reasons for this increase and 120 pick “more work, more business.” Use techniques presented in this chapter and an alpha of .05 to test to determine whether the 43% U.S. figure for this reason holds true in California.

16.83 Is the number of children that a college student has independent of the type of college or university being attended? Suppose students were randomly selected from three types of colleges and universities and the data shown represent the results of a survey of those students. Use a chi-square test of independence of answer the question. Let $\alpha = .05$.

<table>
<thead>
<tr>
<th>Number of Children</th>
<th>Community College</th>
<th>Large University</th>
<th>Small College</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>178</td>
<td>31</td>
</tr>
<tr>
<td>1</td>
<td>49</td>
<td>141</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>54</td>
<td>8</td>
</tr>
<tr>
<td>3 or more</td>
<td>22</td>
<td>14</td>
<td>6</td>
</tr>
</tbody>
</table>

16.84 Should male managers wear a tie during the workday to command respect and demonstrate professionalism? Suppose a measurement scale has been developed that generates a management professionalism score. A random sample of managers in a high-tech industry is selected for the study, some of whom wear ties at work and others of whom do not. One subordinate is selected randomly from each manager’s department and asked to complete the scale on their boss’s professionalism. Analyze the data taken from these independent groups to determine whether the managers with the ties received significantly higher professionalism scores. Let $\alpha = .05$.

<table>
<thead>
<tr>
<th>WITH TIE</th>
<th>WITHOUT TIE</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>22</td>
<td>17</td>
</tr>
</tbody>
</table>

16.85 Many fast-food restaurants have soft-drink dispensers with present amounts, so that when the operator merely pushes a button for the desired drink the cup is automatically filled. This method apparently saves time and

seems to increase worker productivity. To test this conclusion, a researcher randomly selects 18 workers from the fast-food industry, nine from a restaurant with automatic soft-drink dispensers and nine from a comparable restaurant with manual soft-drink dispensers. The samples are independent. During a comparable hour, the amount of sales rung up by the worker is recorded. Assume that $\alpha = .01$ and that a $t$ test is not appropriate. Test whether workers with automatic dispensers are significantly more productive (higher sales per hour).

<table>
<thead>
<tr>
<th>AUTOMATIC DISPENSER</th>
<th>MANUAL DISPENSER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$155$</td>
<td>$105$</td>
</tr>
<tr>
<td>128</td>
<td>118</td>
</tr>
<tr>
<td>143</td>
<td>129</td>
</tr>
<tr>
<td>110</td>
<td>114</td>
</tr>
<tr>
<td>152</td>
<td>125</td>
</tr>
<tr>
<td>168</td>
<td>117</td>
</tr>
<tr>
<td>144</td>
<td>106</td>
</tr>
<tr>
<td>137</td>
<td>92</td>
</tr>
<tr>
<td>118</td>
<td>126</td>
</tr>
</tbody>
</table>

16.86 A particular metal part can be produced at different temperatures. All other variables being equal, a company would like to determine whether the strength of the metal part is significantly different for different temperatures. Given are the strengths of random samples of parts produced under different temperatures. Use $\alpha = .01$ and determine whether there is a significant difference in the strength of the part for different temperatures.

<table>
<thead>
<tr>
<th>TEMPERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>45°</td>
</tr>
<tr>
<td>55°</td>
</tr>
<tr>
<td>70°</td>
</tr>
<tr>
<td>85°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
</tr>
<tr>
<td>215</td>
</tr>
<tr>
<td>218</td>
</tr>
<tr>
<td>216</td>
</tr>
<tr>
<td>219</td>
</tr>
<tr>
<td>220</td>
</tr>
<tr>
<td>221</td>
</tr>
<tr>
<td>223</td>
</tr>
<tr>
<td>224</td>
</tr>
<tr>
<td>225</td>
</tr>
</tbody>
</table>

16.87 Is there a strong correlation between the number of miles driven by a salesperson and sales volume achieved? Data were gathered from nine salespeople who worked territories of similar size and potential. Determine the correlation coefficient for these data. Assume the data are ordinal in level of measurement.

<table>
<thead>
<tr>
<th>SALES</th>
<th>MILES PER MONTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>$150,000$</td>
<td>1500</td>
</tr>
<tr>
<td>$210,000$</td>
<td>2100</td>
</tr>
<tr>
<td>$285,000$</td>
<td>3200</td>
</tr>
<tr>
<td>$301,000$</td>
<td>2400</td>
</tr>
<tr>
<td>$335,000$</td>
<td>2200</td>
</tr>
<tr>
<td>$390,000$</td>
<td>2500</td>
</tr>
<tr>
<td>$400,000$</td>
<td>3300</td>
</tr>
<tr>
<td>$425,000$</td>
<td>3100</td>
</tr>
<tr>
<td>$440,000$</td>
<td>3600</td>
</tr>
</tbody>
</table>

16.88 Workers in three different but comparable companies were asked to rate the use of quality control techniques
in their firms on a 50-point scale. A score of 50 represents nearly perfect implementation of quality control techniques and 0 represents no implementation. Workers are divided into three independent groups. One group worked in a company that had required all its workers to attend a 3-day seminar on quality control 1 year ago. A second group worked in a company in which each worker was part of a quality circle group that had been meeting at least once a month for a year. The third group of workers was employed by a company in which management had been actively involved in the quality control process for more than a year. Use $\alpha = 0.10$ to determine whether there is a significant difference between the three groups, as measured by the ratings.

<table>
<thead>
<tr>
<th>ATTENDED 3-DAY SEMINAR</th>
<th>QUALITY CIRCLES</th>
<th>MANAGEMENT INVOLVED</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>38</td>
<td>21</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>28</td>
</tr>
<tr>
<td>22</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>31</td>
</tr>
</tbody>
</table>

16.89 The scores given are husband-wife scores on a marketing measure. Use the Wilcoxon matched-pairs signed rank test to determine whether the wives’ scores are significantly higher on the marketing measure than the husbands’. Assume that $\alpha = 0.01$.

<table>
<thead>
<tr>
<th>HUSBANDS</th>
<th>WIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>25</td>
<td>29</td>
</tr>
<tr>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>33</td>
</tr>
<tr>
<td>25</td>
<td>38</td>
</tr>
<tr>
<td>22</td>
<td>34</td>
</tr>
<tr>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>23</td>
<td>36</td>
</tr>
<tr>
<td>30</td>
<td>31</td>
</tr>
</tbody>
</table>

16.90 Study the following MINITAB output. What statistical test was run? What type of design was it? What was the result of the test?

Friedman Test
Friedman test of Observations by Treatment blocked by Block
$S = 11.31$ $DF = 3$ $P = 0.010$
$S = 12.16$ $DF = 3$ $P = 0.007$ (adjusted for ties)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Median</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>20.125</td>
<td>17.0</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>25.875</td>
<td>33.0</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>24.500</td>
<td>30.5</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>22.500</td>
<td>19.5</td>
</tr>
</tbody>
</table>

Grand median = 23.250

16.91 Examine the following MINITAB output. Discuss the statistical test, its intent, and its outcome.

Runs Test
$K = 1.4200$
The observed number of runs = 28
The expected number of runs = 25.3600
21 Observations above $K 29 below
The test is significant at 0.4387
Cannot reject an alpha $= 0.05$

16.92 Study the following MINITAB output. What statistical test was being computed by MINITAB? What are the results of this analysis?

Mann-Whitney Confidence Interval and Test
C1 $N = 16$ Median $= 37.000$
C2 $N = 16$ Median $= 46.500$
Point estimate for ETA1-ETA2 is $-8.000$
95.2 Percent C.I. For ETA1-ETA2 is $(-13.999, -2.997)$
$W = 191.5$
Test of ETA1 = ETA2 vs. ETA1 $\neq$ ETA2 is significant at 0.0067
The test is significant at 0.0066 (adjusted for ties)

16.93 Study the following MINITAB output. What type of statistical test was done? What were the hypotheses and what was the outcome? Discuss.

Kruskal-Wallis Test
Kruskal-Wallis test on Observations

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Median</th>
<th>Ave Rank</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>35.00</td>
<td>14.8</td>
<td>0.82</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>25.50</td>
<td>4.2</td>
<td>-3.33</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>35.00</td>
<td>15.0</td>
<td>1.11</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>35.00</td>
<td>16.0</td>
<td>1.40</td>
</tr>
<tr>
<td>Overall</td>
<td>24</td>
<td></td>
<td>12.5</td>
<td></td>
</tr>
</tbody>
</table>

$H = 11.21$ $DF = 3$ $P = 0.011$
$H = 11.28$ $DF = 3$ $P = 0.010$ (adjusted for ties)
SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

17.17 Create a flowchart from the following sequence of activities: Begin. Flow to activity A. Flow to decision B. If yes, flow to activity C. If no, flow to activity D. From C flow to activity E and to activity F. From F, flow to decision G. If yes, flow to decision H. If no at G, stop. At H, if yes, flow to activity I and on to activity J and then stop. If no at H, flow to activity J and stop. At D, flow to activity K, flow to L, and flow to decision M. If yes at M, stop. If no at M, flow to activity N, then stop.

17.18 An examination of rejects shows that there are at least 10 problems. A frequency tally of the problems follows. Construct a Pareto chart for these data.

<table>
<thead>
<tr>
<th>PROBLEM</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>673</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>108</td>
</tr>
<tr>
<td>4</td>
<td>379</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>564</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>402</td>
</tr>
<tr>
<td>9</td>
<td>54</td>
</tr>
<tr>
<td>10</td>
<td>202</td>
</tr>
</tbody>
</table>

17.19 A brainstorm session on possible causes of a problem resulted in five possible causes: A, B, C, D, and E. Cause A has three possible subcauses, cause B has four, cause C has two, cause D has five, and cause E has three. Construct a fishbone diagram for this problem and its possible causes.

17.20 Solve the following.

a. A random sample of 13 items is taken from a lot. If fewer than two items are defective, the lot will be accepted. If two or more are defective, the lot will be rejected. Suppose the lot contains 5% defective items, and that is acceptable to the customer. What is the probability the lot will be rejected? What is the probability it will be accepted? Suppose the lot contains 12% defective items, and that is unacceptable to the customer. What is the probability that the customer will incorrectly accept the lot?

b. A lot has 8575 items. The customer plans to sample 20 of these items randomly to determine whether or not to accept the lot. If no more than two of the sample are defective, the customer will accept the lot. Suppose the lot contains only 3% defective items, which is quite acceptable to the customer. What is the producer's risk?
17.21 A bottled-water company has been randomly inspecting bottles of water to determine whether they are acceptable for delivery and sale. The inspectors are looking at water quality, bottle condition, and seal tightness. A series of 10 random samples of 50 bottles each has been taken. Some bottles have been rejected. Use the following information on the number of bottles from each batch that were rejected as being out of compliance to construct a $P$ chart.

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>$N$</th>
<th>NUMBER OUT OF COMPLIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>6</td>
</tr>
</tbody>
</table>

17.22 A fruit juice company sells a glass container filled with 24 ounces of cranapple juice. Inspectors are concerned about the consistency of volume of fill in these containers. Every 2 hours for 3 days of production, a sample of five containers is randomly selected and the volume of fill is measured. The results follow.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
<th>SAMPLE 3</th>
<th>SAMPLE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.05</td>
<td>24.01</td>
<td>24.03</td>
<td>23.98</td>
</tr>
<tr>
<td>24.01</td>
<td>24.02</td>
<td>23.95</td>
<td>24.00</td>
</tr>
<tr>
<td>24.02</td>
<td>24.10</td>
<td>24.00</td>
<td>24.01</td>
</tr>
<tr>
<td>23.99</td>
<td>24.03</td>
<td>24.01</td>
<td>24.01</td>
</tr>
<tr>
<td>24.04</td>
<td>24.08</td>
<td>23.99</td>
<td>24.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAMPLE 5</th>
<th>SAMPLE 6</th>
<th>SAMPLE 7</th>
<th>SAMPLE 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.97</td>
<td>24.02</td>
<td>24.01</td>
<td>24.08</td>
</tr>
<tr>
<td>23.99</td>
<td>24.05</td>
<td>24.00</td>
<td>24.03</td>
</tr>
<tr>
<td>24.02</td>
<td>24.01</td>
<td>24.00</td>
<td>24.00</td>
</tr>
<tr>
<td>24.01</td>
<td>24.00</td>
<td>23.97</td>
<td>24.05</td>
</tr>
<tr>
<td>24.00</td>
<td>24.01</td>
<td>24.02</td>
<td>24.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAMPLE 9</th>
<th>SAMPLE 10</th>
<th>SAMPLE 11</th>
<th>SAMPLE 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.00</td>
<td>24.00</td>
<td>24.01</td>
<td>24.00</td>
</tr>
<tr>
<td>24.02</td>
<td>24.01</td>
<td>23.99</td>
<td>24.05</td>
</tr>
<tr>
<td>24.03</td>
<td>24.00</td>
<td>24.02</td>
<td>24.04</td>
</tr>
<tr>
<td>24.01</td>
<td>24.00</td>
<td>24.03</td>
<td>24.02</td>
</tr>
<tr>
<td>24.01</td>
<td>24.00</td>
<td>24.01</td>
<td>24.00</td>
</tr>
</tbody>
</table>

17.23 A motor company purchases industrial hoses in lots of 500. The company uses acceptance sampling to determine whether it will accept or reject the hoses. It uses a single-sample plan with $n = 15$. The company will not accept a lot unless there are no nonconforming hoses. Construct an OC curve for this situation. Suppose 2% of the hoses in a lot are nonconforming and this is acceptable to the motor company. What is the producer's risk? Suppose the lot contains .10 that are in nonconformance, which is unacceptable. What is the consumer's risk?

17.24 A metal-manufacturing company produces sheet metal. Statistical quality control technicians randomly select sheets to be inspected for blemishes and size problems. The number of nonconformances per sheet is tallied. Shown here are the results of testing 36 sheets of metal. Use the data to construct a $c$ chart. What is the centerline? What is the meaning of the centerline value?

<table>
<thead>
<tr>
<th>SHEET NUMBER</th>
<th>NUMBER OF NONCONFORMANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SHEET NUMBER</th>
<th>NUMBER OF NONCONFORMANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
</tr>
<tr>
<td>36</td>
<td>3</td>
</tr>
</tbody>
</table>

17.25 A manufacturing company produces cylindrical tubes for engines that are specified to be 1.20 cm thick. As part of the company's statistical quality control effort, random samples of four tubes are taken each hour. The tubes are measured to determine whether they are within thickness tolerances. Shown here are the thickness data in centimeters for nine samples of tubes. Use these data to develop an $X$ chart and an $R$ chart. Comment on whether or not the process appears to be in control at this point.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
<th>SAMPLE 3</th>
<th>SAMPLE 4</th>
<th>SAMPLE 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.22</td>
<td>1.20</td>
<td>1.21</td>
<td>1.16</td>
<td>1.24</td>
</tr>
<tr>
<td>1.19</td>
<td>1.20</td>
<td>1.18</td>
<td>1.17</td>
<td>1.20</td>
</tr>
<tr>
<td>1.20</td>
<td>1.22</td>
<td>1.17</td>
<td>1.20</td>
<td>1.21</td>
</tr>
<tr>
<td>1.23</td>
<td>1.20</td>
<td>1.20</td>
<td>1.16</td>
<td>1.18</td>
</tr>
</tbody>
</table>
17.26 A manufacturer produces digital watches. Every 2 hours a sample of six watches is selected randomly to be tested. Each watch is run for exactly 15 minutes and is timed by an accurate, precise timing device. Because of the variation among watches, they do not all run the same. Shown here are the data from eight different samples given in minutes. Use these data to construct $X$ and $R$ charts. Observe the results and comment on whether or not the process is in control.

<table>
<thead>
<tr>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
<th>SAMPLE 3</th>
<th>SAMPLE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.01</td>
<td>15.03</td>
<td>14.96</td>
<td>15.00</td>
</tr>
<tr>
<td>14.99</td>
<td>14.96</td>
<td>14.97</td>
<td>15.01</td>
</tr>
<tr>
<td>14.99</td>
<td>15.01</td>
<td>14.96</td>
<td>14.97</td>
</tr>
<tr>
<td>15.00</td>
<td>15.02</td>
<td>14.99</td>
<td>15.01</td>
</tr>
<tr>
<td>14.98</td>
<td>14.97</td>
<td>15.01</td>
<td>14.99</td>
</tr>
<tr>
<td>14.99</td>
<td>15.01</td>
<td>14.98</td>
<td>14.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAMPLE 5</th>
<th>SAMPLE 6</th>
<th>SAMPLE 7</th>
<th>SAMPLE 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.02</td>
<td>15.02</td>
<td>15.03</td>
<td>14.96</td>
</tr>
<tr>
<td>15.03</td>
<td>15.01</td>
<td>15.04</td>
<td>14.99</td>
</tr>
<tr>
<td>14.99</td>
<td>14.97</td>
<td>15.03</td>
<td>15.02</td>
</tr>
<tr>
<td>15.01</td>
<td>15.00</td>
<td>15.00</td>
<td>15.01</td>
</tr>
<tr>
<td>15.02</td>
<td>15.01</td>
<td>15.01</td>
<td>14.98</td>
</tr>
<tr>
<td>15.01</td>
<td>14.99</td>
<td>14.99</td>
<td>15.02</td>
</tr>
</tbody>
</table>

17.27 A company produces outdoor home thermometers. For a variety of reasons, a thermometer can be tested and found to be out of compliance with company specification. The company takes samples of thermometers on a regular basis and tests each one to determine whether it meets company standards. Shown here are data from 12 different random samples of 75 thermometers. Use these data to construct a $P$ chart. Comment on the pattern of points in the chart.

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>$n$</th>
<th>NUMBER OUT OF COMPLIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>75</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>75</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>75</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>75</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>75</td>
<td>7</td>
</tr>
</tbody>
</table>

17.28 A plastics company makes thousands of plastic bottles for another company that manufactures saline solution for users of soft contact lenses. The plastics company randomly inspects a sample of its bottles as part of its quality control program. Inspectors look for blemishes on the bottle, size and thickness, closability, leaks, labeling problems, and so on. Shown here are the results of tests completed on 25 bottles. Use these data to construct a $c$ chart. Observe the results and comment on the chart.

<table>
<thead>
<tr>
<th>BOTTLE NUMBER</th>
<th>NUMBER OF NONCONFORMANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BOTTLE NUMBER</th>
<th>NUMBER OF NONCONFORMANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

17.29 A manufacturer of men's jeans purchases zippers in lots of 800. The jeans manufacturer uses single-sample acceptance sampling with a sample size of 10 to determine whether or not to accept the lot. The manufacturer uses $c = 2$ as the acceptance number. Construct an $OC$ curve for this problem. Suppose the lot actually contains 10% nonconforming zippers and this is acceptable to the manufacturer. What is the producer's risk? Suppose the lot contains 30% unacceptable zippers and this is not acceptable. What is the consumer's risk? (Hint: Use Table A.2 to work this problem.)

17.30 A bathtub manufacturer closely inspects several tubs on every shift for nonconformances such as leaks, lack of symmetry, unstable base, drain malfunctions, and so on. The following list gives the number of nonconformances per tub for 40 tubs. Use these data to construct a $c$ chart of nonconformances for bathtubs. Comment on the results of this chart.

<table>
<thead>
<tr>
<th>NUMBER OF TUB NONCONFORMANCES</th>
<th>NUMBER OF TUB NONCONFORMANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
</tr>
</tbody>
</table>

Continued
17.31 A glass manufacturer produces hand mirrors. Each mirror is supposed to meet company standards for such things as glass thickness, reflectivity, size of handle, quality of glass, color of handle, and so on. To control for these features, the company quality people randomly sample 40 mirrors every shift and determine how many of the mirrors are out of compliance on at least one feature. Shown here are the data for 15 such samples. Use the data to construct a P chart. Observe the results and comment on the control of the process as indicated by the chart.

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>n</th>
<th>NUMBER OUT OF COMPLIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>

**INTERPRETING THE OUTPUT**

17.32 Study the Minitab \( \bar{X} \) chart on the fill of a product that is supposed to contain 12 ounces. Does the process appear to be out of control? Why or why not?

17.33 Study the Minitab \( R \) chart for the product and data used in Problem 17.32. Comment on the state of the production process for this item.

17.34 Study the Minitab \( P \) chart for a manufactured item. The chart represents the results of testing 30 items at a time for compliance. Sixty different samples were taken for this chart. Discuss the results and the implications for the production process.

17.35 Study the Minitab \( c \) chart for nonconformances for a part produced in a manufacturing process. Comment on the results.
1. A dairy company in the manufacturing database has been testing its quart milk container fills for volume in four-container samples. Shown here are the results of 10 such samples and the volume measurements in quarts. Use the information to construct both an $\bar{X}$ and an $R$ chart for the data. Discuss the results. What are the centerlines, LCL, and UCL for each of these charts?

<table>
<thead>
<tr>
<th>SAMPLE NUMBER</th>
<th>MEASUREMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.98</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
</tr>
<tr>
<td>3</td>
<td>1.11</td>
</tr>
<tr>
<td>4</td>
<td>.95</td>
</tr>
<tr>
<td>5</td>
<td>1.03</td>
</tr>
<tr>
<td>6</td>
<td>1.04</td>
</tr>
<tr>
<td>7</td>
<td>.94</td>
</tr>
<tr>
<td>8</td>
<td>1.03</td>
</tr>
<tr>
<td>9</td>
<td>1.01</td>
</tr>
<tr>
<td>10</td>
<td>1.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUMBER OF NONCONFORMING STATEMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMPLE 1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

3. A company in the financial database is manufacturing laptop computers. A supplier is shipping the company a particular part in lots of 5000. The computer manufacturing company (consumer) is doing acceptance sampling with samples of size 25. If the supplier is actually producing lots with only 3% defective parts and if this is acceptable to the consumer, what is the producer’s risk if $c = 1$? Suppose the supplier is actually producing the part in lots with 12% defective parts, which is unacceptable to the consumer. If $c$ is still 1, what is the consumer’s risk? Construct an OC curve for $n = 25$ and $c = 1$. Comment on the curve.

2. A hospital in the hospital database takes weekly samples of patient account statements for 12 weeks with each sample containing 40 accounts. Auditors analyze the account statements, looking for nonconforming statements. Shown here are the results of the 12 samples. Use these data to construct a $P$ chart for proportion of nonconforming statements. What is the centerline? What are UCL and LCL? Comment on the control chart.
CALCULATING THE STATISTICS

18.16 Use the following decision table to complete parts (a) through (d).

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>( s_1 )</th>
<th>( s_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>-75</td>
<td>200</td>
</tr>
<tr>
<td>( d_3 )</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>( d_4 )</td>
<td>75</td>
<td>10</td>
</tr>
</tbody>
</table>

a. Use the maximax criterion to determine which decision alternative to select.
b. Use the maximin criterion to determine which decision alternative to select.
c. Use the Hurwicz criterion to determine which decision alternative to select. Let \( \alpha = .6 \).
d. Compute an opportunity loss table from these data. Use this table and a minimax regret criterion to determine which decision alternative to select.

18.17 Use the following decision table to complete parts (a) through (c).

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>( s_1(.30) )</th>
<th>( s_2(.25) )</th>
<th>( s_3(.20) )</th>
<th>( s_4(.25) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>400</td>
<td>250</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>300</td>
<td>-100</td>
<td>600</td>
<td>200</td>
</tr>
</tbody>
</table>

a. Draw a decision tree to represent this decision table.

b. Compute the expected monetary values for each decision and label the decision tree to indicate what the final decision would be.
c. Compute the expected payoff of perfect information. Compare this answer to the answer determined in part (b) and compute the value of perfect information.

18.18 Shown here is a decision table. A forecast can be purchased by the decision maker. The forecaster is not correct 100% of the time. Also given is a table containing the probabilities of the forecast being correct under different states of nature. Use the first table to compute the expected monetary value of this decision without sample information. Use the second table to revise the prior probabilities of the various decision alternatives. From this and the first table, compute the expected monetary value with sample information. Construct a decision tree to represent the options, the payoffs, and the expected monetary values. Calculate the value of sample information.

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>( s_1(.40) )</th>
<th>( s_2(.60) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>$200</td>
<td>$150</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>-$75</td>
<td>$450</td>
</tr>
<tr>
<td>( d_3 )</td>
<td>$175</td>
<td>$125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Forecast</th>
<th>( s_1 )</th>
<th>( s_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>.90</td>
<td>.30</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>.10</td>
<td>.70</td>
</tr>
</tbody>
</table>
18.19 Managers of a manufacturing firm have decided to add Christmas tree ornaments to their list of production items. However, they have not decided how many to produce because they are uncertain about the level of demand. Shown here is a decision table that has been constructed to help the managers in their decision situation. Use this table to answer parts (a) through (c).

<table>
<thead>
<tr>
<th>State of Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
</tr>
<tr>
<td>Small Number</td>
</tr>
<tr>
<td>Modest Number</td>
</tr>
<tr>
<td>Large Number</td>
</tr>
</tbody>
</table>

- Use maximax and maximin criteria to evaluate the decision alternatives.
- Construct an opportunity loss table and use minimax regret to select a decision alternative.
- Compare the results of the maximax, maximin, and minimax regret criteria in selecting decision alternatives.

18.20 Some companies have used production learning curves to set pricing strategies. They price their product lower than the initial cost of making the product; after some period of time, the learning curve takes effect and the product can be produced for less than its selling price. In this way, the company can penetrate new markets with aggressive pricing strategies and still make a long-term profit.

A company is considering using the learning curve to set its price on a new product. There is some uncertainty as to how soon, if at all, the production operation will learn to make the product more quickly and efficiently. If the learning curve does not drop enough and/or the initial price is too low, the company will be operating at a loss on this product. If the product is priced too high, the sales volume might be too low to justify production. Shown here is a decision table that contains as its states of nature several possible learning-curve scenarios. The decision alternatives are three different pricing strategies. Use this table and the Hurwitz criterion to make a decision about the pricing strategies with each given value of alpha.

<table>
<thead>
<tr>
<th>State of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Learning</td>
</tr>
<tr>
<td>Price Low</td>
</tr>
<tr>
<td>Price Medium</td>
</tr>
<tr>
<td>Price High</td>
</tr>
</tbody>
</table>

- $\alpha = .10$
- $\alpha = .50$
- $\alpha = .80$
- Compare and discuss the decision choices in parts (a) through (c).

18.21 An entertainment company owns two amusement parks in the South. They are faced with the decision of whether or not to open the parks in the winter. If they choose to open the parks in the winter, they can leave the parks open during regular hours (as in the summer) or they can open only on the weekends. To some extent, the payoffs from opening the park hinge on the type of weather that occurs during the winter season. Following are the payoffs for various decision options about opening the park for two different weather scenarios: mild weather and severe weather. Use the information to construct a decision tree. Determine the expected monetary value and the value of perfect information.

<table>
<thead>
<tr>
<th>State of the Weather</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
</tr>
<tr>
<td>(.75)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decision</th>
<th>Open Regular Hours</th>
<th>Open Weekends Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative</td>
<td>$2000</td>
<td>$1200</td>
</tr>
<tr>
<td></td>
<td>$-2500</td>
<td>$-200</td>
</tr>
<tr>
<td></td>
<td>$-300</td>
<td>$100</td>
</tr>
</tbody>
</table>

18.22 A U.S. manufacturing company has decided to consider producing a particular model of one of its products just for sale in Germany. Because of the German requirements, the product must be made specifically for German consumption and cannot be sold in the United States. Company officials believe the market for the product is very price-sensitive. As the product will be manufactured in the United States and exported to Germany, the biggest variable factor in being price competitive is the exchange rate between the two countries. If the dollar is strong, German consumers will have to pay more for the product in marks. If the dollar becomes weaker against the mark, Germans can buy more U.S. products for their money. The company officials are faced with decision alternatives of whether to produce the product or not. The states of the exchange rates are: dollar weaker, dollar stays the same, and dollar stronger. The probabilities of these states occurring are .35, .25, and .40, respectively. There are some negative payoffs for not producing the product because of sunk development and market research costs and because of lost market opportunity. If the product is not produced, the payoffs are $-700$ when the dollar gets weaker, $-200$ when the dollar remains about the same, and $150$ when the dollar gets stronger. If the product is produced, the payoffs are $1800$ when the dollar gets weaker, $400$ when the exchange rates stay about the same, and $-1600$ when the dollar gets stronger.

Use this information to construct a decision tree and a decision table for this decision-making situation. Use the probabilities to compute the expected monetary values of the decision alternatives. On the basis of this information, which decision choice should the company make? Compare the expected monetary value of perfect information and the value of perfect information.
18.23 a. A small retailer began as a mom-and-pop operation selling crafts and consignment items. In the past 2 years, the store's volume has grown significantly. The owners are trying to decide whether or not to purchase an automated checkout system. Their present manual system is slow. They are concerned about lost business due to inability to ring up sales quickly. The automated system would also have some accounting and inventory advantages. The problem is that the automated system carries a large fixed cost, and the owners feel that sales volume would have to grow to justify the cost.

The following decision table contains the decision alternatives for this situation, the possible states of future sales, prior probabilities of those states occurring, and the payoffs. Use this information to compute the expected monetary payoffs for the alternatives.

<table>
<thead>
<tr>
<th>State of Sales</th>
<th>Reduction</th>
<th>Constant</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automate</td>
<td>-$40,000</td>
<td>-$15,000</td>
<td>$60,000</td>
</tr>
<tr>
<td>Don't Automate</td>
<td>$5,000</td>
<td>$10,000</td>
<td>-$30,000</td>
</tr>
</tbody>
</table>

b. For a fee, the owners can purchase a sales forecast for the near future. The forecast is not always perfect. The probabilities of these forecasts being correct for particular states of sales are shown here. Use these probabilities to revise the prior state probabilities. Compute the expected monetary value on the basis of sample information. Determine the value of the sample information.

<table>
<thead>
<tr>
<th>State of Sales</th>
<th>Reduction</th>
<th>Constant</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction</td>
<td>.60</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>Constant</td>
<td>.30</td>
<td>.80</td>
<td>.25</td>
</tr>
<tr>
<td>Increase</td>
<td>.10</td>
<td>.10</td>
<td>.70</td>
</tr>
</tbody>
</table>

18.24 a. A city is considering airport expansion. In particular, the mayor and city council are trying to decide whether or not to sell bonds to construct a new terminal. The problem is that there is not enough demand for gates at present to warrant construction of a new terminal. However, a major airline is investigating several cities to determine which will become its new hub. If this city is selected, the new terminal will easily pay for itself. The decision to build the terminal must be made by the city before the airline will say whether or not the city has been chosen as its hub. Shown here is a decision table for this dilemma. Use this information to compute expected monetary values for the alternatives and reach a conclusion.

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>City Chosen (0.20)</th>
<th>City Not Chosen (0.80)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td>Build Terminal</td>
<td>Don't Build Terminal</td>
</tr>
<tr>
<td>Alternative</td>
<td>$12,000</td>
<td>-$8,000</td>
</tr>
<tr>
<td></td>
<td>-$1,000</td>
<td>$2,000</td>
</tr>
</tbody>
</table>

b. An airline industry expert indicates that she will sell the city decision makers her best “guess” as to whether or not the city will be chosen as hub for the airline. The probabilities of her being right or wrong are given. Use these probabilities to revise the prior probabilities of the city being chosen as the hub and then calculate the expected monetary value by using the sample information. Determine the value of the sample information for this problem.

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>City Chosen</th>
<th>City Not Chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City Chosen</td>
<td>.45</td>
<td>.40</td>
</tr>
<tr>
<td>City Not Chosen</td>
<td>.55</td>
<td>.60</td>
</tr>
</tbody>
</table>