Chapter 1
Data and Statistics

Solutions:

2. a. The ten elements are the ten tablet computers

   b. 5 variables: Cost ($), Operating System, Display Size (inches), Battery Life (hours), CPU Manufacturer

   c. Categorical variables: Operating System and CPU Manufacturer

       Quantitative variables: Cost ($), Display Size (inches), and Battery Life (hours)

   d.

       | Variable              | Measurement Scale |
       |-----------------------|-------------------|
       | Cost ($)              | Ratio             |
       | Operating System      | Nominal           |
       | Display Size (inches) | Ratio             |
       | Battery Life (hours)  | Ratio             |
       | CPU Manufacturer      | Nominal           |

4. a. There are eight elements in this data set; each element corresponds to one of the eight models of cordless telephones

   b. Categorical variables: Voice Quality and Handset on Base

       Quantitative variables: Price, Overall Score, and Talk Time

   c. Price – ratio measurement

       Overall Score – interval measurement

       Voice Quality – ordinal measurement

       Handset on Base – nominal measurement

       Talk Time – ratio measurement

6. a. Categorical

   b. Quantitative

   c. Categorical

   d. Quantitative

   e. Quantitative
8. a. 762
   b. Categorical
   c. Percentages
   d. \( \cdot67(762) = 510.54 \)

   510 or 511 respondents said they want the amendment to pass.

10. a. Categorical
    b. Percentages
    c. 44 of 1080 respondents or approximately 4% strongly agree with allowing drivers of motor vehicles to talk on a hand-held cell phone while driving.
    d. 165 of the 1080 respondents or 15% of said they somewhat disagree and 741 or 69% said they strongly disagree. Thus, there does not appear to be general support for allowing drivers of motor vehicles to talk on a hand-held cell phone while driving.

12. a. The population is all visitors coming to the state of Hawaii.
    b. Since airline flights carry the vast majority of visitors to the state, the use of questionnaires for passengers during incoming flights is a good way to reach this population. The questionnaire actually appears on the back of a mandatory plants and animals declaration form that passengers must complete during the incoming flight. A large percentage of passengers complete the visitor information questionnaire.
    c. Questions 1 and 4 provide quantitative data indicating the number of visits and the number of days in Hawaii. Questions 2 and 3 provide categorical data indicating the categories of reason for the trip and where the visitor plans to stay.

13. a. Google revenue in billions of dollars
    b. Quantitative
    c. Time series
    d. Google revenue is increasing over time.
14. a. The graph of the time series follows:

![Time Series Graph]

b. In Year 1 and Year 2 Hertz was the clear market share leader. In Year 3 and Year 4 Hertz and Avis have approximately the same market share. The market share for Dollar appears to be declining.

c. The bar chart for Year 4 is shown below.

![Bar Chart]

This chart is based on cross-sectional data.
16. The answer to this exercise depends on updating the time series of the average price per gallon of conventional regular gasoline as shown in Figure 1.1. Contact the website www.eia.doe.gov to obtain the most recent time series data. The answer should focus on the most recent changes or trend in the average price per gallon.

18. a. 684/1021; or approximately 67%
   b. (.6)*(1021) = 612.6  Therefore, 612 or 613 used an accountant or professional tax preparer.
   c. Categorical

20. a. 43% of managers were bullish or very bullish.
   21% of managers expected health care to be the leading industry over the next 12 months.
   b. We estimate the average 12-month return estimate for the population of investment managers to be 11.2%.
   c. We estimate the average over the population of investment managers to be 2.5 years.

22. a. The population consists of all clients that currently have a home listed for sale with the agency or have hired the agency to help them locate a new home.
   b. Some of the ways that could be used to collect the data are as follows:
   • A survey could be mailed to each of the agency’s clients.
   • Each client could be sent an email with a survey attached.
   • The next time one of the firm’s agents meets with a client they could conduct a personal interview to obtain the data.

24. a. This is a statistically correct descriptive statistic for the sample.
   b. An incorrect generalization since the data was not collected for the entire population.
   c. An acceptable statistical inference based on the use of the word “estimate.”
   d. While this statement is true for the sample, it is not a justifiable conclusion for the entire population.
   e. This statement is not statistically supportable. While it is true for the particular sample observed, it is entirely possible and even very likely that at least some students will be outside the 65 to 90 range of grades.
Chapter 2
Descriptive Statistics: Tabular and Graphical Displays

Solutions:

2.  a.  $1 - (0.22 + 0.18 + 0.40) = 0.20$

   b.  $0.20(200) = 40$

   c/d.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$0.22(200) = 44$</td>
<td>22</td>
</tr>
<tr>
<td>B</td>
<td>$0.18(200) = 36$</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>$0.40(200) = 80$</td>
<td>40</td>
</tr>
<tr>
<td>D</td>
<td>$0.20(200) = 40$</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100</td>
</tr>
</tbody>
</table>

3.  a.  $360\degree \times \frac{58}{120} = 174\degree$

   b.  $360\degree \times \frac{42}{120} = 126\degree$

   c.  

   ![Pie Chart]

   No Opinion 16.7%
   No 35.0%
   Yes 48.3%
d.

4. a. These data are categorical.

b.

<table>
<thead>
<tr>
<th>Show</th>
<th>Frequency</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jep</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>JJ</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>OWS</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>THM</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>WoF</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

c.

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
d. The largest viewing audience is for *Wheel of Fortune* and the second largest is for *Two and a Half Men*.

6. a.  

<table>
<thead>
<tr>
<th>Network</th>
<th>Relative Frequency</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>CBS</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>FOX</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>NBC</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>25</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

b. For these data, NBC and CBS tie for the number of top-rated shows. Each has 9 (36%) of the top 25. ABC is third with 6 (24%) and the much younger FOX network has 1 (4%).
7. a. 

<table>
<thead>
<tr>
<th>Rating</th>
<th>Frequency</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Very Good</td>
<td>23</td>
<td>46</td>
</tr>
<tr>
<td>Good</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Fair</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

\[ \frac{50}{100} \]

Management should be very pleased with the survey results. 40% + 46% = 86% of the ratings are very good to excellent. 94% of the ratings are good or better. This does not look to be a Delta flight where significant changes are needed to improve the overall customer satisfaction ratings.

b. While the overall ratings look fine, note that one customer (2%) rated the overall experience with the flight as Fair and two customers (4%) rated the overall experience with the flight as Poor. It might be insightful for the manager to review explanations from these customers as to how the flight failed to meet expectations. Perhaps, it was an experience with other passengers that Delta could do little to correct or perhaps it was an isolated incident that Delta could take steps to correct in the future.

8. a. 

<table>
<thead>
<tr>
<th>Position</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitcher</td>
<td>17</td>
<td>0.309</td>
</tr>
<tr>
<td>Catcher</td>
<td>4</td>
<td>0.073</td>
</tr>
<tr>
<td>1st Base</td>
<td>5</td>
<td>0.091</td>
</tr>
<tr>
<td>2nd Base</td>
<td>4</td>
<td>0.073</td>
</tr>
<tr>
<td>3rd Base</td>
<td>2</td>
<td>0.036</td>
</tr>
<tr>
<td>Shortstop</td>
<td>5</td>
<td>0.091</td>
</tr>
<tr>
<td>Left Field</td>
<td>6</td>
<td>0.109</td>
</tr>
<tr>
<td>Center Field</td>
<td>5</td>
<td>0.091</td>
</tr>
<tr>
<td>Right Field</td>
<td>7</td>
<td>0.127</td>
</tr>
</tbody>
</table>

\[ 55 \]

b. Pitchers (Almost 31%)

c. 3rd Base (3 – 4%)

d. Right Field (Almost 13%)
e. Infielders (16 or 29.1%) to Outfielders (18 or 32.7%)

10. a.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>187</td>
</tr>
<tr>
<td>Very Good</td>
<td>252</td>
</tr>
<tr>
<td>Average</td>
<td>107</td>
</tr>
<tr>
<td>Poor</td>
<td>62</td>
</tr>
<tr>
<td>Terrible</td>
<td>41</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>649</strong></td>
</tr>
</tbody>
</table>

b.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>29</td>
</tr>
<tr>
<td>Very Good</td>
<td>39</td>
</tr>
<tr>
<td>Average</td>
<td>16</td>
</tr>
<tr>
<td>Poor</td>
<td>10</td>
</tr>
<tr>
<td>Terrible</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

c. ![](chart.png)

d. 29% + 39% = 68% of the guests at the Sheraton Anaheim Hotel rated the hotel as Excellent or Very Good. But, 10% + 6% = 16% of the guests rated the hotel as poor or terrible.

e. The percent frequency distribution for Disney’s Grand Californian follows:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>48</td>
</tr>
<tr>
<td>Very Good</td>
<td>31</td>
</tr>
<tr>
<td>Average</td>
<td>12</td>
</tr>
<tr>
<td>Poor</td>
<td>6</td>
</tr>
</tbody>
</table>
48% + 31% = 79% of the guests at the Sheraton Anaheim Hotel rated the hotel as Excellent or Very Good. And, 6% + 3% = 9% of the guests rated the hotel as poor or terrible.

Compared to ratings of other hotels in the same region, both of these hotels received very favorable ratings. But, in comparing the two hotels, guests at Disney’s Grand Californian provided somewhat better ratings than guests at the Sheraton Anaheim Hotel.

12.

<table>
<thead>
<tr>
<th>Class</th>
<th>Cumulative Frequency</th>
<th>Cumulative Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than or equal to 19</td>
<td>10</td>
<td>.20</td>
</tr>
<tr>
<td>less than or equal to 29</td>
<td>24</td>
<td>.48</td>
</tr>
<tr>
<td>less than or equal to 39</td>
<td>41</td>
<td>.82</td>
</tr>
<tr>
<td>less than or equal to 49</td>
<td>48</td>
<td>.96</td>
</tr>
<tr>
<td>less than or equal to 59</td>
<td>50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

14. a.

b/c.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0 – 7.9</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>8.0 – 9.9</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>10.0 – 11.9</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>12.0 – 13.9</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>14.0 – 15.9</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>100</td>
</tr>
</tbody>
</table>

15. Leaf Unit = .1

<table>
<thead>
<tr>
<th>6</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>5 5 7</td>
</tr>
<tr>
<td>8</td>
<td>1 3 4 8</td>
</tr>
<tr>
<td>9</td>
<td>3 6</td>
</tr>
<tr>
<td>10</td>
<td>0 4 5</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>
16. Leaf Unit = 10

| 11 | 6 |
| 12 | 0 2 |
| 13 | 0 6 7 |
| 14 | 2 2 7 |
| 15 | 5 |
| 16 | 0 2 8 |
| 17 | 0 2 3 |

17. a/b.

<table>
<thead>
<tr>
<th>Waiting Time</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 4</td>
<td>4</td>
<td>0.20</td>
</tr>
<tr>
<td>5 – 9</td>
<td>8</td>
<td>0.40</td>
</tr>
<tr>
<td>10 – 14</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>15 – 19</td>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>20 – 24</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Totals</td>
<td>20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

c/d.

<table>
<thead>
<tr>
<th>Waiting Time</th>
<th>Cumulative Frequency</th>
<th>Cumulative Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than or equal to 4</td>
<td>4</td>
<td>0.20</td>
</tr>
<tr>
<td>Less than or equal to 9</td>
<td>12</td>
<td>0.60</td>
</tr>
<tr>
<td>Less than or equal to 14</td>
<td>17</td>
<td>0.85</td>
</tr>
<tr>
<td>Less than or equal to 19</td>
<td>19</td>
<td>0.95</td>
</tr>
<tr>
<td>Less than or equal to 24</td>
<td>20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

e. 12/20 = 0.60

18. a.

<table>
<thead>
<tr>
<th>PPG</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-11.9</td>
<td>1</td>
</tr>
<tr>
<td>12-13.9</td>
<td>3</td>
</tr>
<tr>
<td>14-15.9</td>
<td>7</td>
</tr>
<tr>
<td>16-17.9</td>
<td>19</td>
</tr>
<tr>
<td>18-19.9</td>
<td>9</td>
</tr>
<tr>
<td>20-21.9</td>
<td>4</td>
</tr>
<tr>
<td>22-23.9</td>
<td>2</td>
</tr>
<tr>
<td>24-25.9</td>
<td>0</td>
</tr>
<tr>
<td>26-27.9</td>
<td>3</td>
</tr>
<tr>
<td>28-29.9</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>
b.

<table>
<thead>
<tr>
<th>PPG</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-11.9</td>
<td>0.02</td>
</tr>
<tr>
<td>12-13.9</td>
<td>0.06</td>
</tr>
<tr>
<td>14-15.9</td>
<td>0.14</td>
</tr>
<tr>
<td>16-17.9</td>
<td>0.38</td>
</tr>
<tr>
<td>18-19.9</td>
<td>0.18</td>
</tr>
<tr>
<td>20-21.9</td>
<td>0.08</td>
</tr>
<tr>
<td>22-23.9</td>
<td>0.04</td>
</tr>
<tr>
<td>24-25.9</td>
<td>0.00</td>
</tr>
<tr>
<td>26-27.9</td>
<td>0.06</td>
</tr>
<tr>
<td>28-29.9</td>
<td>0.04</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
</tr>
</tbody>
</table>

c.

<table>
<thead>
<tr>
<th>PPG</th>
<th>Cumulative Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than 12</td>
<td>2</td>
</tr>
<tr>
<td>less than 14</td>
<td>8</td>
</tr>
<tr>
<td>less than 16</td>
<td>22</td>
</tr>
<tr>
<td>less than 18</td>
<td>60</td>
</tr>
<tr>
<td>less than 20</td>
<td>78</td>
</tr>
<tr>
<td>less than 22</td>
<td>86</td>
</tr>
<tr>
<td>less than 24</td>
<td>90</td>
</tr>
<tr>
<td>less than 26</td>
<td>90</td>
</tr>
<tr>
<td>less than 28</td>
<td>96</td>
</tr>
<tr>
<td>less than 30</td>
<td>100</td>
</tr>
</tbody>
</table>

d.

e. There is skewness to the right.
20. a. Lowest = 12, Highest = 23

b. | Hours in Meetings per Week | Frequency | Percent Frequency |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11-12</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>13-14</td>
<td>2</td>
<td>8%</td>
</tr>
<tr>
<td>15-16</td>
<td>6</td>
<td>24%</td>
</tr>
<tr>
<td>17-18</td>
<td>3</td>
<td>12%</td>
</tr>
<tr>
<td>19-20</td>
<td>5</td>
<td>20%</td>
</tr>
<tr>
<td>21-22</td>
<td>4</td>
<td>16%</td>
</tr>
<tr>
<td>23-24</td>
<td>4</td>
<td>16%</td>
</tr>
</tbody>
</table>

The distribution is slightly skewed to the left.

c.

<table>
<thead>
<tr>
<th># U.S. Locations</th>
<th>Frequency</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4999</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>5000-9999</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>10000-14999</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>15000-19999</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>20000-24999</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25000-29999</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>30000-34999</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>35000-39999</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Total: 20 100
b. The distribution is skewed to the right. The majority of the franchises in this list have fewer than 20,000 locations (50% + 15% + 15% = 80%). McDonald's, Subway and 7-Eleven have the highest number of locations.

24. Leaf Unit = 1000

<table>
<thead>
<tr>
<th>Starting Median Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

Leaf Unit = 1000

<table>
<thead>
<tr>
<th>Mid-Career Median Salary</th>
</tr>
</thead>
<tbody>
<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>

There is a wider spread in the mid-career median salaries than in the starting median salaries. Also, as expected, the mid-career median salaries are higher than the starting median salaries. The mid-career median salaries were mostly in the $93,000 to $114,000 range while the starting median salaries were mostly in the $51,000 to $62,000 range.
26. a.

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

b. Most frequent age group: 40-44 with 9 runners
c. 43 was the most frequent age with 5 runners

27. a.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>11</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>

b.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>100.0</td>
<td>84.6</td>
<td>16.7</td>
<td>100.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>15.4</td>
<td>83.3</td>
<td>100.0</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>

c.
### Chapter 2

#### Table 2.1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>27.8</td>
<td>0.0</td>
</tr>
<tr>
<td>B</td>
<td>61.1</td>
<td>16.7</td>
</tr>
<tr>
<td>C</td>
<td>11.1</td>
<td>83.3</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

#### d. Category A values for $x$ are always associated with category 1 values for $y$. Category B values for $x$ are usually associated with category 1 values for $y$. Category C values for $x$ are usually associated with category 2 values for $y$. 

#### 28. a.

<table>
<thead>
<tr>
<th></th>
<th>20-39</th>
<th>40-59</th>
<th>60-79</th>
<th>80-100</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-29</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>x 30-49</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>50-69</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>70-90</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Grand Total</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>

#### b.

<table>
<thead>
<tr>
<th></th>
<th>20-39</th>
<th>40-59</th>
<th>60-79</th>
<th>80-100</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-29</td>
<td>20.0</td>
<td>80.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>x 30-49</td>
<td>33.3</td>
<td>66.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>50-69</td>
<td>20.0</td>
<td>60.0</td>
<td>20.0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>70-90</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Grand Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

#### c.

<table>
<thead>
<tr>
<th></th>
<th>20-39</th>
<th>40-59</th>
<th>60-79</th>
<th>80-100</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-29</td>
<td>0.0</td>
<td>0.0</td>
<td>16.7</td>
<td>100.0</td>
<td>100</td>
</tr>
<tr>
<td>x 30-49</td>
<td>28.6</td>
<td>0.0</td>
<td>66.7</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>50-69</td>
<td>14.3</td>
<td>100.0</td>
<td>16.7</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>70-90</td>
<td>57.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>Grand Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

#### d. Higher values of $x$ are associated with lower values of $y$ and vice versa.
30. a. Row Percentages

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Speed</td>
<td>130-139.9</td>
<td>140-149.9</td>
<td>150-159.9</td>
<td>160-169.9</td>
<td>170-179.9</td>
<td></td>
</tr>
<tr>
<td>130-139.9</td>
<td>16.7</td>
<td>0.0</td>
<td>0.0</td>
<td>50.0</td>
<td>25.0</td>
<td>100</td>
</tr>
<tr>
<td>140-149.9</td>
<td>25.0</td>
<td>25.0</td>
<td>25.0</td>
<td>12.5</td>
<td>12.5</td>
<td>100</td>
</tr>
<tr>
<td>150-159.9</td>
<td>0.0</td>
<td>50.0</td>
<td>16.7</td>
<td>16.7</td>
<td>16.7</td>
<td>100</td>
</tr>
<tr>
<td>160-169.9</td>
<td>50.0</td>
<td>0.0</td>
<td>50.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>170-179.9</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
</tr>
</tbody>
</table>

b. It appears that most of the faster average winning times occur before 2003. This could be due to new regulations that take into account driver safety, fan safety, the environmental impact, and fuel consumption during races.

32. a. Row percentages are shown below.

<table>
<thead>
<tr>
<th>Region</th>
<th>Under $15,000</th>
<th>$15,000 to $24,999</th>
<th>$25,000 to $34,999</th>
<th>$35,000 to $49,999</th>
<th>$50,000 to $74,999</th>
<th>$75,000 to $99,999</th>
<th>$100,000 and over</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>12.72</td>
<td>10.45</td>
<td>10.54</td>
<td>13.07</td>
<td>17.22</td>
<td>11.57</td>
<td>24.42</td>
<td>100.00</td>
</tr>
<tr>
<td>Midwest</td>
<td>12.40</td>
<td>12.60</td>
<td>11.58</td>
<td>14.27</td>
<td>19.11</td>
<td>12.06</td>
<td>17.97</td>
<td>100.00</td>
</tr>
<tr>
<td>South</td>
<td>14.30</td>
<td>12.97</td>
<td>11.55</td>
<td>14.85</td>
<td>17.73</td>
<td>11.04</td>
<td>17.57</td>
<td>100.00</td>
</tr>
<tr>
<td>West</td>
<td>11.84</td>
<td>10.73</td>
<td>10.15</td>
<td>13.65</td>
<td>18.44</td>
<td>11.77</td>
<td>23.43</td>
<td>100.00</td>
</tr>
</tbody>
</table>

The percent frequency distributions for each region now appear in each row of the table. For example, the percent frequency distribution of the West region is as follows:

<table>
<thead>
<tr>
<th>Income Level</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under $15,000</td>
<td>11.84</td>
</tr>
<tr>
<td>$15,000 to $24,999</td>
<td>10.73</td>
</tr>
<tr>
<td>$25,000 to $34,999</td>
<td>10.15</td>
</tr>
<tr>
<td>$35,000 to $49,999</td>
<td>13.65</td>
</tr>
<tr>
<td>$50,000 to $74,999</td>
<td>18.44</td>
</tr>
<tr>
<td>$75,000 to $99,999</td>
<td>11.77</td>
</tr>
<tr>
<td>$100,000 and over</td>
<td>23.43</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
</tr>
</tbody>
</table>

b. West: 18.44 + 11.77 + 23.43 = 53.64% or \( \frac{4804 + 3066 + 6104}{26057} = 53.63\% \)

South: 17.73 + 11.04 + 17.57 = 46.34% or \( \frac{7730 + 4813 + 7660}{43609} = 46.33\% \)

c.
Chapter 2

Northeast

Percent Frequency vs. Income Level

Midwest

Percent Frequency vs. Income Level

South

Percent Frequency vs. Income Level

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
The largest difference appears to be a higher percentage of household incomes of $100,000 and over for the Northeast and West regions.

d. Column percentages are shown below.

<table>
<thead>
<tr>
<th>Region</th>
<th>Under $15,000</th>
<th>$15,000 to $24,999</th>
<th>$25,000 to $34,999</th>
<th>$35,000 to $49,999</th>
<th>$50,000 to $74,999</th>
<th>$75,000 to $99,999</th>
<th>$100,000 and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>17.83</td>
<td>16.00</td>
<td>17.41</td>
<td>16.90</td>
<td>17.38</td>
<td>18.35</td>
<td>22.09</td>
</tr>
<tr>
<td>Midwest</td>
<td>21.35</td>
<td>23.72</td>
<td>23.50</td>
<td>22.68</td>
<td>23.71</td>
<td>23.49</td>
<td>19.96</td>
</tr>
<tr>
<td>South</td>
<td>40.68</td>
<td>40.34</td>
<td>38.75</td>
<td>39.00</td>
<td>36.33</td>
<td>35.53</td>
<td>32.25</td>
</tr>
<tr>
<td>West</td>
<td>20.13</td>
<td>19.94</td>
<td>20.34</td>
<td>21.42</td>
<td>22.58</td>
<td>22.63</td>
<td>25.70</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Each column is a percent frequency distribution of the region variable for one of the household income categories. For example, for an income level of $35,000 to $49,999 the percent frequency distribution for the region variable is as follows:

<table>
<thead>
<tr>
<th>Region</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>16.90</td>
</tr>
<tr>
<td>Midwest</td>
<td>22.68</td>
</tr>
<tr>
<td>South</td>
<td>39.00</td>
</tr>
<tr>
<td>West</td>
<td>21.42</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
</tr>
</tbody>
</table>

e. 32.25% of households with a household income of $100,000 and over are from the South, while 17.57% of households from the South have income of $100,000 and over. These percentages are different because they represent percent frequencies based on different category totals.
34. a. 

<table>
<thead>
<tr>
<th>Industry</th>
<th>0-25</th>
<th>25-50</th>
<th>50-75</th>
<th>75-100</th>
<th>100-125</th>
<th>125-150</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automotive &amp; Luxury</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Consumer Packaged Goods</td>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Financial Services</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>Technology</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>41</td>
<td>14</td>
<td>10</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>82</td>
</tr>
</tbody>
</table>

b. 

<table>
<thead>
<tr>
<th>Brand Revenue ($ billions)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-25</td>
<td>41</td>
</tr>
<tr>
<td>25-50</td>
<td>14</td>
</tr>
<tr>
<td>50-75</td>
<td>10</td>
</tr>
<tr>
<td>75-100</td>
<td>5</td>
</tr>
<tr>
<td>100-125</td>
<td>7</td>
</tr>
<tr>
<td>125-150</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>82</td>
</tr>
</tbody>
</table>

c. Consumer packaged goods have the lowest brand revenues; each of the 12 consumer packaged goods brands in the sample data had a brand revenue of less than $25 billion. Approximately 57% of the financial services brands (8 out of 14) had a brand revenue of $50 billion or greater, and 47% of the technology brands (7 out of 15) had a brand revenue of at least $50 billion.

d. 

<table>
<thead>
<tr>
<th>Industry</th>
<th>-60--41</th>
<th>-40--21</th>
<th>-20--1</th>
<th>0-19</th>
<th>20-39</th>
<th>40-60</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automotive &amp; Luxury</td>
<td>11</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Consumer Packaged Goods</td>
<td>2</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Financial Services</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>20</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Technology</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1</td>
<td>4</td>
<td>14</td>
<td>52</td>
<td>10</td>
<td>1</td>
<td>82</td>
</tr>
</tbody>
</table>

e. 

<table>
<thead>
<tr>
<th>1-Yr Value Change (%)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-60--41</td>
<td>1</td>
</tr>
<tr>
<td>-40--21</td>
<td>4</td>
</tr>
<tr>
<td>-20--1</td>
<td>14</td>
</tr>
</tbody>
</table>
f. The automotive & luxury brands all had a positive 1-year value change (%). The technology brands had the greatest variability. Financial services were heavily concentrated between -20 and +19 % changes, while consumer goods and other industries were mostly concentrated in 0-19% gains.

36. a.

b. There is a negative relationship between $x$ and $y$; $y$ decreases as $x$ increases.

38. a.

<table>
<thead>
<tr>
<th>$x$</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>66.667</td>
<td>30.000</td>
<td>80.000</td>
</tr>
<tr>
<td>No</td>
<td>33.333</td>
<td>70.000</td>
<td>20.000</td>
</tr>
</tbody>
</table>

b.
40. a.

b. Colder average low temperature seems to lead to higher amounts of snowfall.

c. Two cities have an average snowfall of nearly 100 inches of snowfall: Buffalo, N.Y and Rochester, NY. Both are located near large lakes in New York.
42. a. After an increase in age 25-34, smartphone ownership decreases as age increases. The percentage of people with no cell phone increases with age. There is less variation across age groups in the percentage who own other cell phones.

c. Unless a newer device replaces the smartphone, we would expect smartphone ownership would become less sensitive to age. This would be true because current users will become older and because the device will become to be seen more as a necessity than a luxury.

44. a.  

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>800-999</td>
<td>1</td>
</tr>
<tr>
<td>1000-1199</td>
<td>3</td>
</tr>
<tr>
<td>1200-1399</td>
<td>6</td>
</tr>
<tr>
<td>1400-1599</td>
<td>10</td>
</tr>
<tr>
<td>1600-1799</td>
<td>7</td>
</tr>
<tr>
<td>1800-1999</td>
<td>2</td>
</tr>
<tr>
<td>2000-2199</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
</tr>
</tbody>
</table>
b. The distribution is nearly symmetrical. It could be approximated by a bell-shaped curve.

c. 10 of 30 or 33% of the scores are between 1400 and 1599. The average SAT score looks to be a little over 1500. Scores below 800 or above 2200 are unusual.

46. a.

<table>
<thead>
<tr>
<th>Population in Millions</th>
<th>Frequency</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 2.4</td>
<td>15</td>
<td>30.0%</td>
</tr>
<tr>
<td>2.5-4.9</td>
<td>13</td>
<td>26.0%</td>
</tr>
<tr>
<td>5.0-7.4</td>
<td>10</td>
<td>20.0%</td>
</tr>
<tr>
<td>7.5-9.9</td>
<td>5</td>
<td>10.0%</td>
</tr>
<tr>
<td>10.0-12.4</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>12.5-14.9</td>
<td>2</td>
<td>4.0%</td>
</tr>
<tr>
<td>15.0-17.4</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>17.5-19.9</td>
<td>2</td>
<td>4.0%</td>
</tr>
<tr>
<td>20.0-22.4</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>22.5-24.9</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>25.0-27.4</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>27.5-29.9</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>30.0-32.4</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>32.5-34.9</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>35.0-37.4</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>37.5-39.9</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
b. The distribution is skewed to the right.

c. 15 states (30%) have a population less than 2.5 million. Over half of the states have population less than 5 million (28 states – 56%). Only seven states have a population greater than 10 million (California, Florida, Illinois, New York, Ohio, Pennsylvania and Texas). The largest state is California (37.3 million) and the smallest states are Vermont and Wyoming (600 thousand).

48. a.

<table>
<thead>
<tr>
<th>Industry</th>
<th>Frequency</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank</td>
<td>26</td>
<td>13%</td>
</tr>
<tr>
<td>Cable</td>
<td>44</td>
<td>22%</td>
</tr>
<tr>
<td>Car</td>
<td>42</td>
<td>21%</td>
</tr>
<tr>
<td>Cell</td>
<td>60</td>
<td>30%</td>
</tr>
<tr>
<td>Collection</td>
<td>28</td>
<td>14%</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>100%</td>
</tr>
</tbody>
</table>

b.
c. The cellular phone providers had the highest number of complaints.

d. The percentage frequency distribution shows that the two financial industries (banks and collection agencies) had about the same number of complaints. Also, new car dealers and cable and satellite television companies also had about the same number of complaints.

50.a.

<table>
<thead>
<tr>
<th>Level of Education</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School graduate</td>
<td>32,773/65,644(100) = 49.93</td>
</tr>
<tr>
<td>Bachelor's degree</td>
<td>22,131/65,644(100) = 33.71</td>
</tr>
<tr>
<td>Master's degree</td>
<td>9003/65,644(100) = 13.71</td>
</tr>
<tr>
<td>Doctoral degree</td>
<td>1737/65,644(100) = 2.65</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
</tr>
</tbody>
</table>

13.71 + 2.65 = 16.36% of heads of households have a master’s or doctoral degree.

b.

<table>
<thead>
<tr>
<th>Household Income</th>
<th>Percent Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under $25,000</td>
<td>13,128/65,644(100) = 20.00</td>
</tr>
<tr>
<td>$25,000 to $49,999</td>
<td>15,499/65,644(100) = 23.61</td>
</tr>
<tr>
<td>$50,000 to $99,999</td>
<td>20,548/65,644(100) = 31.30</td>
</tr>
<tr>
<td>$100,000 and over</td>
<td>16,469/65,644(100) = 25.09</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
</tr>
</tbody>
</table>

31.30 + 25.09 = 56.39% of households have an income of $50,000 or more.

c.

<table>
<thead>
<tr>
<th>Level of Education</th>
<th>Household Income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under $25,000</td>
</tr>
<tr>
<td>High School graduate</td>
<td></td>
</tr>
<tr>
<td>Bachelor's degree</td>
<td></td>
</tr>
<tr>
<td>Master's degree</td>
<td></td>
</tr>
<tr>
<td>Doctoral degree</td>
<td></td>
</tr>
</tbody>
</table>
There is a large difference between the level of education for households with an income of under $25,000 and households with an income of $100,000 or more. For instance, 75.26% of households with an income of under $25,000 are households in which the head of the household is a high school graduate. But, only 21.14% of households with an income level of $100,000 or more are households in which the head of the household is a high school graduate. It is interesting to note, however, that 45.95% of households with an income of $50,000 to $99,999 are households in which the head of the household is a high school graduate.

52 a.

<table>
<thead>
<tr>
<th>Size of Company</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10-0</td>
<td>12</td>
</tr>
<tr>
<td>0-10</td>
<td>60</td>
</tr>
<tr>
<td>10-20</td>
<td>13</td>
</tr>
<tr>
<td>20-30</td>
<td>3</td>
</tr>
<tr>
<td>30-40</td>
<td>0</td>
</tr>
<tr>
<td>60-70</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>98</td>
</tr>
</tbody>
</table>

b. Frequency distribution for growth rate.

<table>
<thead>
<tr>
<th>Job Growth (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10-0</td>
<td>12</td>
</tr>
<tr>
<td>0-10</td>
<td>60</td>
</tr>
<tr>
<td>10-20</td>
<td>13</td>
</tr>
<tr>
<td>20-30</td>
<td>8</td>
</tr>
<tr>
<td>30-40</td>
<td>4</td>
</tr>
<tr>
<td>60-70</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>98</td>
</tr>
</tbody>
</table>

Frequency distribution for size of company.

<table>
<thead>
<tr>
<th>Size</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>32</td>
</tr>
<tr>
<td>Medium</td>
<td>28</td>
</tr>
<tr>
<td>Large</td>
<td>38</td>
</tr>
</tbody>
</table>
c. Crosstabulation showing column percentages.

<table>
<thead>
<tr>
<th>Size of Company</th>
<th>Job Growth (%)</th>
<th>Small</th>
<th>Midsized</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-10-0</td>
<td>13</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0-10</td>
<td>56</td>
<td>46</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>10-20</td>
<td>22</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>20-30</td>
<td>9</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>30-40</td>
<td>0</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>60-70</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

d. Crosstabulation showing row percentages.

<table>
<thead>
<tr>
<th>Size of Company</th>
<th>Job Growth (%)</th>
<th>Small</th>
<th>Midsized</th>
<th>Large</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-10-0</td>
<td>33</td>
<td>50</td>
<td>17</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0-10</td>
<td>30</td>
<td>22</td>
<td>48</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>10-20</td>
<td>54</td>
<td>15</td>
<td>31</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>20-30</td>
<td>38</td>
<td>38</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>30-40</td>
<td>0</td>
<td>75</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>60-70</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

e. 12 companies had a negative job growth: 13% were small companies; 21% were midsized companies; and 5% were large companies. So, in terms of avoiding negative job growth, large companies were better off than small and midsized companies. But, although 95% of the large companies had a positive job growth, the growth rate was below 10% for 76% of these companies. In terms of better job growth rates, midsized companies performed better than either small or large companies. For instance, 26% of the midsized companies had a job growth of at least 20% as compared to 9% for small companies and 8% for large companies.

54. a.  

<table>
<thead>
<tr>
<th>Year Founded</th>
<th>35-40</th>
<th>45-55</th>
<th>65-75</th>
<th>85-95</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600-1649</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1700-1749</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1750-1799</td>
<td></td>
<td>1</td>
<td>3</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>1800-1849</td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1850-1899</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1900-1949</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1950-2000</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>103</td>
</tr>
</tbody>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
b. Older colleges and universities tend to have higher graduation rates.

<table>
<thead>
<tr>
<th>Year Founded</th>
<th>35-40</th>
<th>40-45</th>
<th>45-50</th>
<th>50-55</th>
<th>55-60</th>
<th>60-65</th>
<th>65-70</th>
<th>70-75</th>
<th>75-80</th>
<th>80-85</th>
<th>85-90</th>
<th>90-95</th>
<th>95-100</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600-1649</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1700-1749</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1750-1799</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75.00</td>
<td>100</td>
</tr>
<tr>
<td>1800-1849</td>
<td>4.76</td>
<td>9.52</td>
<td>19.05</td>
<td></td>
<td>9.52</td>
<td>14.29</td>
<td></td>
<td>19.05</td>
<td>14.29</td>
<td></td>
<td></td>
<td></td>
<td>9.52</td>
<td>100</td>
</tr>
<tr>
<td>1850-1899</td>
<td>2.04</td>
<td>4.08</td>
<td>8.16</td>
<td>6.12</td>
<td>22.45</td>
<td>10.20</td>
<td>18.37</td>
<td>12.24</td>
<td>6.12</td>
<td>8.16</td>
<td>2.04</td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1900-1949</td>
<td>5.56</td>
<td>5.56</td>
<td>5.56</td>
<td>5.56</td>
<td>16.67</td>
<td>16.67</td>
<td>11.11</td>
<td>22.22</td>
<td>5.56</td>
<td>5.56</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1950-2000</td>
<td>14.29</td>
<td>14.29</td>
<td>42.86</td>
<td>28.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

c. There appears to be a strong positive relationship between Tuition & Fees and % Graduation.

b. There appears to be a strong positive relationship between Tuition & Fees and % Graduation.

58. a.
Zoo attendance appears to be dropping over time.

b. 

c. General attendance is increasing, but not enough to offset the decrease in member attendance. School membership appears fairly stable.
Chapter 3
Descriptive Statistics: Numerical Measures

Solutions:

2. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{96}{6} = 16 \]
   
   10, 12, 16, 17, 20, 21
   
   Median = \[ \frac{16 + 17}{2} = 16.5 \]

3. a. \[ \bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{6(3.2) + 3(2) + 2(2.5) + 8(5)}{6 + 3 + 2 + 8} = \frac{70.2}{19} = 3.69 \]
   
   b. \[ \frac{3.2 + 2 + 2.5 + 5}{4} = \frac{12.7}{4} = 3.175 \]

4. 

<table>
<thead>
<tr>
<th>Period</th>
<th>Rate of Return (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-6.0</td>
</tr>
<tr>
<td>2</td>
<td>-8.0</td>
</tr>
<tr>
<td>3</td>
<td>-4.0</td>
</tr>
<tr>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>5.4</td>
</tr>
</tbody>
</table>

The mean growth factor over the five periods is:

\[ \bar{x} = \sqrt[5]{(x_1)(x_2) \cdots (x_5)} = \sqrt[5]{(0.940)(0.920)(0.960)(1.020)(1.054)} = \sqrt[5]{0.8925} = 0.9775 \]

So the mean growth rate \((0.9775 - 1)100\% = -2.25\%\).

5. 15, 20, 25, 25, 27, 28, 30, 34

\[ L_{20} = \frac{p}{100} (n + 1) = \frac{20}{100} (8 + 1) = 1.8 \]

20th percentile = 15 + .8(20 - 15) = 19

\[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (8 + 1) = 2.25 \]

25th percentile = 20 + .25(25 - 20) = 21.25
Chapter 3

\[ L_{65} = \frac{p}{100} (n + 1) = \frac{65}{100} (8 + 1) = 5.85 \]

65th percentile = \( 27 + .85(28 – 27) = 27.85 \)

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (8 + 1) = 6.75 \]

75th percentile = \( 28 + .75(30 – 28) = 29.5 \)

6. \[ \text{Mean} = \frac{\sum x_i}{n} = \frac{657}{11} = 59.73 \]

Median = 57 6th item

Mode = 53 It appears 3 times

8. a. Median = 80 or $80,000. The median salary for the sample of 15 middle-level managers working at firms in Atlanta is slightly lower than the median salary reported by the Wall Street Journal.

b. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{1260}{15} = 84 \]

Mean salary is $84,000. The sample mean salary for the sample of 15 middle-level managers is greater than the median salary. This indicates that the distribution of salaries for middle-level managers working at firms in Atlanta is positively skewed.

c. The sorted data are as follows:

53 55 63 67 73 75 77 80 83 85 93 106 108 118 124

\[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (16) = 4 \]

First quartile or 25th percentile is the value in position 4 or 67.

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (16) = 12 \]

Third quartile or 75th percentile is the value in position 12 or 106.

10. a. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{1318}{20} = 65.9 \]

Order the data from the lowest rating (42) to the highest rating (83)

<table>
<thead>
<tr>
<th>Position</th>
<th>Rating</th>
<th>Position</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>12</td>
<td>67</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>13</td>
<td>68</td>
</tr>
<tr>
<td>4</td>
<td>61</td>
<td>14</td>
<td>69</td>
</tr>
<tr>
<td>5</td>
<td>61</td>
<td>15</td>
<td>71</td>
</tr>
</tbody>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
Descriptive Statistics: Numerical Measures

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>61</td>
<td>16</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>17</td>
<td>76</td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>18</td>
<td>78</td>
</tr>
<tr>
<td>9</td>
<td>64</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>66</td>
<td>20</td>
<td>83</td>
</tr>
</tbody>
</table>

\[ L_{50} = \frac{p}{100} (n + 1) = \frac{50}{100} (20 + 1) = 10.5 \]
Median or 50th percentile = 66 + .5(67 – 66) = 66.5
Mode is 61.

b. \[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (20 + 1) = 5.25 \]
First quartile or 25th percentile = 61

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (20 + 1) = 15.75 \]
Third quartile or 75th percentile = 71

c. \[ L_{90} = \frac{p}{100} (n + 1) = \frac{90}{100} (20 + 1) = 18.9 \]
90th percentile = 78 + .9(81 – 78) = 80.7
90% of the ratings are 80.7 or less; 10% of the ratings are 80.7 or greater.

12. a. The minimum number of viewers that watched a new episode is 13.3 million, and the maximum number is 16.5 million.

b. The mean number of viewers that watched a new episode is 15.04 million or approximately 15.0 million; the median also 15.0 million. The data is multimodal (13.6, 14.0, 16.1, and 16.2 million); in such cases the mode is usually not reported.

c. The data are first arranged in ascending order.

\[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (21 + 1) = 5.50 \]
First quartile or 25th percentile = 14 + .50(14.1 – 14) = 14.05

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (21 + 1) = 16.5 \]
Third quartile or 75th percentile = 16 + .5(16.1 – 16) = 16.05

d. A graph showing the viewership data over the air dates follows. Period 1 corresponds to the first episode of the season, period 2 corresponds to the second episode, and so on.
This graph shows that viewership of *The Big Bang Theory* has been relatively stable over the 2011–2012 television season.

14. For March 2011:

\[ L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(50+1) = 12.75 \]

First quartile or 25th percentile = 6.8 + .75(6.8 – 6.8) = 6.8

\[ L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(50+1) = 25.5 \]

Second quartile or median = 8 + .5(8 – 8) = 8

\[ L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(50+1) = 38.25 \]

Third quartile or 75th percentile = 9.4 + .25(9.6 – 9.4) = 9.45

For March 2012:

\[ L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(50+1) = 12.75 \]

First quartile or 25th percentile = 6.2 + .75(6.2 – 6.2) = 6.2

\[ L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(50+1) = 25.5 \]

Second quartile or median = 7.3 + .5(7.4 – 7.3) = 7.35

\[ L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(50+1) = 38.25 \]

Third quartile or 75th percentile = 8.6 + .25(8.6 – 8.6) = 8.6
It may be easier to compare these results if we place them in a table.

<table>
<thead>
<tr>
<th>First Quartile</th>
<th>March 2011</th>
<th>March 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.80</td>
<td>6.20</td>
</tr>
<tr>
<td>Median</td>
<td>8.00</td>
<td>7.35</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>9.45</td>
<td>8.60</td>
</tr>
</tbody>
</table>

The results show that in March 2012 approximately 25% of the states had an unemployment rate of 6.2% or less, lower than in March 2011. And, the median of 7.35% and the third quartile of 8.6% in March 2012 are both less than the corresponding values in March 2011, indicating that unemployment rates across the states are decreasing.

16. a.

<table>
<thead>
<tr>
<th>Grade $x_i$</th>
<th>Weight $w_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 (A)</td>
<td>9</td>
</tr>
<tr>
<td>3 (B)</td>
<td>15</td>
</tr>
<tr>
<td>2 (C)</td>
<td>33</td>
</tr>
<tr>
<td>1 (D)</td>
<td>3</td>
</tr>
<tr>
<td>0 (F)</td>
<td>0</td>
</tr>
</tbody>
</table>

60 Credit Hours

$$
\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{9(4) + 15(3) + 33(2) + 3(1)}{9 + 15 + 33 + 3} = \frac{150}{60} = 2.50
$$

b. Yes; satisfies the 2.5 grade point average requirement

18.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Deans</th>
<th>$w_i x_i$</th>
<th>Recruiters</th>
<th>$w_i x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>44</td>
<td>220</td>
<td>31</td>
<td>155</td>
</tr>
<tr>
<td>4</td>
<td>66</td>
<td>264</td>
<td>34</td>
<td>136</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>180</td>
<td>43</td>
<td>129</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>20</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>684</td>
<td>120</td>
<td>444</td>
</tr>
</tbody>
</table>

Deans: $$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{684}{180} = 3.8$$

Recruiters: $$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{444}{120} = 3.7$$

Business school deans rated the overall academic quality of master’s programs slightly higher than corporate recruiters did.

20.

<table>
<thead>
<tr>
<th>Year</th>
<th>Stivers End of Year Value</th>
<th>Growth Factor</th>
<th>Trippi End of Year Value</th>
<th>Growth Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year1</td>
<td>$11,000</td>
<td>1.100</td>
<td>$5,600</td>
<td>1.120</td>
</tr>
</tbody>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
<table>
<thead>
<tr>
<th>Year</th>
<th>Initial Investment</th>
<th>Growth Rate</th>
<th>Final Value</th>
<th>Annual Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$12,000</td>
<td>1.091</td>
<td>$6,300</td>
<td>1.125</td>
</tr>
<tr>
<td>3</td>
<td>$13,000</td>
<td>1.083</td>
<td>$6,900</td>
<td>1.095</td>
</tr>
<tr>
<td>4</td>
<td>$14,000</td>
<td>1.077</td>
<td>$7,600</td>
<td>1.101</td>
</tr>
<tr>
<td>5</td>
<td>$15,000</td>
<td>1.071</td>
<td>$8,500</td>
<td>1.118</td>
</tr>
<tr>
<td>6</td>
<td>$16,000</td>
<td>1.067</td>
<td>$9,200</td>
<td>1.082</td>
</tr>
<tr>
<td>7</td>
<td>$17,000</td>
<td>1.063</td>
<td>$9,900</td>
<td>1.076</td>
</tr>
<tr>
<td>8</td>
<td>$18,000</td>
<td>1.059</td>
<td>$10,600</td>
<td>1.071</td>
</tr>
</tbody>
</table>

For the Stivers mutual fund we have:

\[18000=10000\left(\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)\right),\] so \[\left(\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)\right)=1.8\] and

\[\bar{x}_g = \sqrt[12]{\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)} = \sqrt[12]{1.80} = 1.07624\]

So the mean annual return for the Stivers mutual fund is \((1.07624 - 1)\times100 = 7.624\%\)

For the Trippi mutual fund we have:

\[10600=5000\left(\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)\right),\] so \[\left(\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)\right)=2.12\] and

\[\bar{x}_g = \sqrt[12]{\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)} = \sqrt[12]{2.12} = 1.09848\]

So the mean annual return for the Trippi mutual fund is \((1.09848 - 1)\times100 = 9.848\%\).

While the Stivers mutual fund has generated a nice annual return of 7.6%, the annual return of 9.8% earned by the Trippi mutual fund is far superior.

22. \[25,000,000=10,000,000\left(\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)\right),\] so \[\left(\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)\right)=2.50,\] and so

\[\bar{x}_g = \sqrt[12]{\left(x_1\right)\left(x_2\right)\cdots\left(x_n\right)} = \sqrt[12]{2.50} = 1.165\]

So the mean annual growth rate is \((1.165 - 1)\times100 = 16.5\%\)

24. \[\bar{x} = \frac{\sum x_i}{n} = \frac{75}{5} = 15\]

\[s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{64}{4} = 16\]

\[s = \sqrt{16} = 4\]

25. Range = 34 – 15 = 19

\[L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (9) = 2.25\]

First Quartile or \(Q_1 = 20 + .25(25-20) = 21.25\)
Descriptive Statistics: Numerical Measures

\[ L_{25} = \frac{p}{100}(n+1) = \frac{75}{100}(9) = 6.75 \]

Third Quartile or \( Q_3 = 28 + .75(30-28) = 29.5 \)

IQR = \( Q_3 - Q_1 = 29.5 - 21.25 = 8.25 \)

\( \bar{x} = \frac{\Sigma x_i}{n} = \frac{204}{8} = 25.5 \)

\( s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = \frac{242}{7} = 34.57 \)

\( s = \sqrt{34.57} = 5.88 \)

26. a. \( \bar{x} = \frac{\Sigma x_i}{n} = \frac{74.4}{20} = 3.72 \)

b. \( s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1.6516}{20-1}} = \sqrt{0.0869} = 0.2948 \)

c. The average price for a gallon of unleaded gasoline in San Francisco is much higher than the national average. This indicates that the cost of living in San Francisco is higher than it would be for cities that have an average price close to the national average.

28. a. The mean serve speed is 180.95, the variance is 21.42, and the standard deviation is 4.63.

b. Although the mean serve speed for the twenty Women's Singles serve speed leaders for the 2011 Wimbledon tournament is slightly higher, the difference is very small. Furthermore, given the variation in the twenty Women's Singles serve speed leaders from the 2012 Australian Open and the twenty Women's Singles serve speed leaders from the 2011 Wimbledon tournament, the difference in the mean serve speeds is most likely due to random variation in the players' performances.

30. Dawson Supply: Range = 11 – 9 = 2

\( s = \sqrt{\frac{4.1}{9}} = 0.67 \)

J.C. Clark: Range = 15 – 7 = 8

\( s = \sqrt{\frac{60.1}{9}} = 2.58 \)

32. a. Automotive: \( \bar{x} = \frac{\Sigma x_i}{n} = \frac{39201}{20} = 1960.05 \)

Department store: \( \bar{x} = \frac{\Sigma x_i}{n} = \frac{13857}{20} = 692.85 \)
b. Automotive:  
\[ s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{4,407,720.95}{19}} = 481.65 \]

Department store:  
\[ s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{456804.55}{19}} = 155.06 \]

c. Automotive:  
\[ 2901 - 598 = 2303 \]
Department Store:  
\[ 1011 - 448 = 563 \]

d. Order the data for each variable from the lowest to highest.

<table>
<thead>
<tr>
<th>Automotive</th>
<th>Department Store</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>598</td>
</tr>
<tr>
<td>2</td>
<td>1512</td>
</tr>
<tr>
<td>3</td>
<td>1573</td>
</tr>
<tr>
<td>4</td>
<td>1642</td>
</tr>
<tr>
<td>5</td>
<td>1714</td>
</tr>
<tr>
<td>6</td>
<td>1720</td>
</tr>
<tr>
<td>7</td>
<td>1781</td>
</tr>
<tr>
<td>8</td>
<td>1798</td>
</tr>
<tr>
<td>9</td>
<td>1813</td>
</tr>
<tr>
<td>10</td>
<td>2008</td>
</tr>
<tr>
<td>11</td>
<td>2014</td>
</tr>
<tr>
<td>12</td>
<td>2024</td>
</tr>
<tr>
<td>13</td>
<td>2058</td>
</tr>
<tr>
<td>14</td>
<td>2166</td>
</tr>
<tr>
<td>15</td>
<td>2202</td>
</tr>
<tr>
<td>16</td>
<td>2254</td>
</tr>
<tr>
<td>17</td>
<td>2366</td>
</tr>
<tr>
<td>18</td>
<td>2526</td>
</tr>
<tr>
<td>19</td>
<td>2531</td>
</tr>
<tr>
<td>20</td>
<td>2901</td>
</tr>
<tr>
<td></td>
<td>448</td>
</tr>
<tr>
<td></td>
<td>472</td>
</tr>
<tr>
<td></td>
<td>474</td>
</tr>
<tr>
<td></td>
<td>573</td>
</tr>
<tr>
<td></td>
<td>589</td>
</tr>
<tr>
<td></td>
<td>597</td>
</tr>
<tr>
<td></td>
<td>598</td>
</tr>
<tr>
<td></td>
<td>622</td>
</tr>
<tr>
<td></td>
<td>629</td>
</tr>
<tr>
<td></td>
<td>669</td>
</tr>
<tr>
<td></td>
<td>706</td>
</tr>
<tr>
<td></td>
<td>714</td>
</tr>
<tr>
<td></td>
<td>746</td>
</tr>
<tr>
<td></td>
<td>760</td>
</tr>
<tr>
<td></td>
<td>782</td>
</tr>
<tr>
<td></td>
<td>824</td>
</tr>
<tr>
<td></td>
<td>840</td>
</tr>
<tr>
<td></td>
<td>856</td>
</tr>
<tr>
<td></td>
<td>947</td>
</tr>
<tr>
<td></td>
<td>1011</td>
</tr>
</tbody>
</table>

\[ L_{25} = \frac{P}{100} (n + 1) = \frac{25}{100} (21) = 5.25 \]

Automotive: First quartile or 25\textsuperscript{th} percentile = 1714 + .25(1720 – 1714) = 1715.5  
Department Store: First quartile or 25\textsuperscript{th} percentile = 589 + .25(597 – 589) = 591

\[ L_{75} = \frac{P}{100} (n + 1) = \frac{75}{100} (21) = 15.75 \]

Automotive: Third quartile or 75\textsuperscript{th} percentile = 2202 + .75(2254 – 2202) = 2241  
Department Store: First quartile or 75\textsuperscript{th} percentile = 782 + .75(824 – 782) = 813.5

Automotive IQR = Q3 – Q1 = 2241 - 1715.5 = 525.5  
Department Store IQR = Q3 – Q1 = 813.5 - 591 = 222.5
e. Automotive spends more on average, has a larger standard deviation, larger max and min, and larger range than Department Store. Autos have all new model years and may spend more heavily on advertising.

34. Quarter milers

\[ s = 0.0564 \]

Coefficient of Variation \(= \frac{s}{\overline{x}} \times 100\% = \frac{0.0564}{0.966} \times 100\% = 5.8\% \)

Milers

\[ s = 0.1295 \]

Coefficient of Variation \(= \frac{s}{\overline{x}} \times 100\% = \frac{0.1295}{4.534} \times 100\% = 2.9\% \)

Yes; the coefficient of variation shows that as a percentage of the mean the quarter milers’ times show more variability.

36. 

\[ z = \frac{520 - 500}{100} = +.20 \]

\[ z = \frac{650 - 500}{100} = +1.50 \]

\[ z = \frac{500 - 500}{100} = 0.00 \]

\[ z = \frac{450 - 500}{100} = -.50 \]

\[ z = \frac{280 - 500}{100} = -2.20 \]

37. a. 

\[ z = \frac{20 - 30}{5} = -2, \quad z = \frac{40 - 30}{5} = 2 \quad 1 - \frac{1}{2^2} = .75 \text{ At least 75\%} \]

b. 

\[ z = \frac{15 - 30}{5} = -3, \quad z = \frac{45 - 30}{5} = 3 \quad 1 - \frac{1}{3^2} = .89 \text{ At least 89\%} \]

c. 

\[ z = \frac{22 - 30}{5} = -1.6, \quad z = \frac{38 - 30}{5} = 1.6 \quad 1 - \frac{1}{1.6^2} = .61 \text{ At least 61\%} \]

d. 

\[ z = \frac{18 - 30}{5} = -2.4, \quad z = \frac{42 - 30}{5} = 2.4 \quad 1 - \frac{1}{2.4^2} = .83 \text{ At least 83\%} \]

e. 

\[ z = \frac{12 - 30}{5} = -3.6, \quad z = \frac{48 - 30}{5} = 3.6 \quad 1 - \frac{1}{3.6^2} = .92 \text{ At least 92\%} \]

38. a. Approximately 95\%

b. Almost all

c. Approximately 68\%
39. a. This is from 2 standard deviations below the mean to 2 standard deviations above the mean.

With \( z = 2 \), Chebyshev’s theorem gives:

\[
1 - \frac{1}{z^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = .75
\]

Therefore, at least 75% of adults sleep between 4.5 and 9.3 hours per day.

b. This is from 2.5 standard deviations below the mean to 2.5 standard deviations above the mean.

With \( z = 2.5 \), Chebyshev’s theorem gives:

\[
1 - \frac{1}{z^2} = 1 - \frac{1}{2.5^2} = 1 - \frac{1}{6.25} = .84
\]

Therefore, at least 84% of adults sleep between 3.9 and 9.9 hours per day.

c. With \( z = 2 \), the empirical rule suggests that 95% of adults sleep between 4.5 and 9.3 hours per day. The percentage obtained using the empirical rule is greater than the percentage obtained using Chebyshev’s theorem.

40. a. \$3.33 is one standard deviation below the mean and \$3.53 is one standard deviation above the mean. The empirical rule says that approximately 68% of gasoline sales are in this price range.

b. Part (a) shows that approximately 68% of the gasoline sales are between \$3.33 and \$3.53. Since the bell-shaped distribution is symmetric, approximately half of 68%, or 34%, of the gasoline sales should be between \$3.33 and the mean price of \$3.43. \$3.63 is two standard deviations above the mean price of \$3.43. The empirical rule says that approximately 95% of the gasoline sales should be within two standard deviations of the mean. Thus, approximately half of 95%, or 47.5%, of the gasoline sales should be between the mean price of \$3.43 and \$3.63. The percentage of gasoline sales between \$3.33 and \$3.63 should be approximately 34% + 47.5% = 81.5%.

c. \$3.63 is two standard deviations above the mean and the empirical rule says that approximately 95% of the gasoline sales should be within two standard deviations of the mean. Thus, \( 1 - .95 = .05 \) of the gasoline sales should be more than two standard deviations from the mean. Since the bell-shaped distribution is symmetric, we expected half of 5%, or 2.5%, would be more than \$3.63.

42. a. \[
z = \frac{x - \mu}{\sigma} = \frac{2300 - 3100}{1200} = -.67
\]

b. \[
z = \frac{x - \mu}{\sigma} = \frac{4900 - 3100}{1200} = 1.50
\]

c. \$2300 is .67 standard deviations below the mean. \$4900 is 1.50 standard deviations above the mean. Neither is an outlier.

d. \[
z = \frac{x - \mu}{\sigma} = \frac{13000 - 3100}{1200} = 8.25
\]

\$13,000 is 8.25 standard deviations above the mean. This cost is an outlier.
44. a. \[ \bar{x} = \frac{\Sigma x_i}{n} = \frac{765}{10} = 76.5 \]

\[ s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{442.5}{10-1}} = 7.01 \]

b. \[ z = \frac{x - \bar{x}}{s} = \frac{84 - 76.5}{7.01} = 1.07 \]

Approximately one standard deviation above the mean. Approximately 68% of the scores are within one standard deviation. Thus, half of the remaining 32%, or 16%, of the games should have a winning score of more than one standard deviation above the mean or a score of 84 or more points.

\[ z = \frac{x - \bar{x}}{s} = \frac{90 - 76.5}{7.01} = 1.93 \]

Approximately two standard deviations above the mean. Approximately 95% of the scores are within two standard deviations. Thus, half of the remaining 5%, or 2.5%, of the games should have a winning score of more than two standard deviations above the mean or a score of more than 90 points.

c. \[ \bar{x} = \frac{\Sigma x_i}{n} = \frac{122}{10} = 12.2 \]

\[ s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{559.6}{10-1}} = 7.89 \]

Smallest margin 3: \[ z = \frac{x - \bar{x}}{s} = \frac{3 - 12.2}{7.89} = -1.17 \]

Largest margin 24: \[ z = \frac{x - \bar{x}}{s} = \frac{24 - 12.2}{7.89} = 1.50 \] No outliers.

46. 15, 20, 25, 25, 27, 28, 30, 34

Smallest = 15

\[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (8 + 1) = 2.25 \]

First quartile or 25th percentile = 20 + .25(25 − 20) = 21.25

\[ L_{50} = \frac{p}{100} (n + 1) = \frac{50}{100} (8 + 1) = 4.5 \]

Second quartile or median = 25 + .5(27 − 25) = 26

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (8 + 1) = 6.75 \]
Third quartile or 75th percentile = 28 + .75(30 − 28) = 29.5
Largest = 34

48.  5, 6, 8, 10, 10, 12, 15, 16, 18
Smallest = 5

\[ L_{25} = \frac{p}{100}(n + 1) = \frac{25}{100}(9 + 1) = 2.5 \]

First quartile or 25th percentile = 6 + .5(8 − 6) = 7

\[ L_{50} = \frac{p}{100}(n + 1) = \frac{50}{100}(9 + 1) = 5.0 \]

Second quartile or median = 10

\[ L_{75} = \frac{p}{100}(n + 1) = \frac{75}{100}(9 + 1) = 7.5 \]

Third quartile or 75th percentile = 15 + .5(16 − 15) = 15.5
Largest = 18

A box plot created using Excel’s Box and Whisker Statistical Chart follows.

50. a. The first place runner in the men’s group finished 109.03 − 65.30 = 43.73 minutes ahead of the first place runner in the women’s group. Lauren Wald would have finished in 11th place for the combined groups.

b. Using Excel’s MEDIAN function the results are as follows:

<table>
<thead>
<tr>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
Using the median finish times, the men’s group finished $131.67 - 109.64 = 22.03$ minutes ahead of the women’s group.

Also note that the fastest time for a woman runner, 109.03 minutes, is approximately equal to the median time of 109.64 minutes for the men’s group.

c. Using Excel’s QUARTILE.EXC function the quartiles are as follows:

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.1025</td>
<td>122.080</td>
</tr>
<tr>
<td>2</td>
<td>109.640</td>
<td>131.670</td>
</tr>
<tr>
<td>3</td>
<td>129.025</td>
<td>147.180</td>
</tr>
</tbody>
</table>

Excel’s MIN and MAX functions provided the following values.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>65.30</td>
<td>109.03</td>
</tr>
<tr>
<td>Maximum</td>
<td>148.70</td>
<td>189.28</td>
</tr>
</tbody>
</table>

Five number summary for men: 65.30, 83.1025, 109.640, 129.025, 148.70

Five number summary for women: 109.03, 122.08, 131.67, 147.18, 189.28

d. Men: $IQR = 129.025 - 83.1025 = 45.9225$

Lower Limit = $Q_1 - 1.5(IQR) = 83.1025 - 1.5(45.9225) = 14.22$

Upper Limit = $Q_3 + 1.5(IQR) = 129.025 + 1.5(45.9225) = 197.91$

There are no outliers in the men’s group.

Women: $IQR = 147.18 - 122.08 = 25.10$

Lower Limit = $Q_1 - 1.5(IQR) = 122.08 - 1.5(25.10) = 84.43$

Upper Limit = $Q_3 + 1.5(IQR) = 147.18 + 1.5(25.10) = 184.83$

The two slowest women runners with times of 189.27 and 189.28 minutes are outliers in the women’s group.
e. A box plot created using Excel’s Box and Whisker Statistical Chart follows.

![Box Plot](image)

The box plots show the men runners with the faster or lower finish times. However, the box plots show the women runners with the lower variation in finish times. The interquartile ranges of 45.9225 minutes for men and 25.10 minutes for women support this conclusion.

51. a. Smallest = 608

\[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (21 + 1) = 5.5 \]

First quartile or 25th percentile = 1850 + .5(1872 – 1850) = 1861

\[ L_{50} = \frac{p}{100} (n + 1) = \frac{50}{100} (21 + 1) = 11.0 \]

Second quartile or median = 4019

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (21 + 1) = 16.5 \]

Third quartile or 75th percentile = 8305 + .5(8408 – 8305) = 8356.5

Largest = 14138

Five-number summary: 608, 1861, 4019, 8356.5, 14138

b. Limits:

\[ IQR = Q_3 - Q_1 = 8356.5 - 1861 = 6495.5 \]

Lower Limit: \( Q_1 - 1.5(IQR) = 1861 - 1.5(6495.5) = -7,882.25 \)

Upper Limit: \( Q_3 + 1.5(IQR) = 8356.5 + 1.5(6495.5) = 18,099.75 \)
c. There are no outliers, all data are within the limits.

d. Yes, if the first two digits in Johnson and Johnson's sales were transposed to 41,138, sales would have shown up as an outlier. A review of the data would have enabled the correction of the data.

e. A box plot created using Excel’s Box and Whisker Statistical Chart follows.

![Box Plot]

52. Excel’s MIN, QUARTILE.EXC, and MAX functions provided the following five-number summaries:

<table>
<thead>
<tr>
<th></th>
<th>AT&amp;T</th>
<th>Sprint</th>
<th>T-Mobile</th>
<th>Verizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>66</td>
<td>63</td>
<td>68</td>
<td>75</td>
</tr>
<tr>
<td>First Quartile</td>
<td>68</td>
<td>65</td>
<td>71.25</td>
<td>77</td>
</tr>
<tr>
<td>Median</td>
<td>71</td>
<td>66</td>
<td>73.5</td>
<td>78.5</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>73</td>
<td>67.75</td>
<td>74.75</td>
<td>79.75</td>
</tr>
<tr>
<td>Maximum</td>
<td>75</td>
<td>69</td>
<td>77</td>
<td>81</td>
</tr>
</tbody>
</table>

a. Median for T-Mobile is 73.5

b. 5- number summary: 68, 71.25, 73.5, 74.75, 77

c. IQR = Q3 – Q1 = 74.75 – 71.25 = 3.5

   Lower Limit = Q1 – 1.5(IQR)

   = 71.25 – 1.5(3.5) = 66

   Upper Limit = Q3 + 1.5(IQR)

   = 74.75 + 1.5(3.5) = 80

   All ratings are between 66 and 80. There are no outliers for the T-Mobile service.
d. Using the five number summaries shown initially, the limits for the four cell-phone services are as follows:

<table>
<thead>
<tr>
<th></th>
<th>AT&amp;T</th>
<th>Sprint</th>
<th>T-Mobile</th>
<th>Verizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>66</td>
<td>63</td>
<td>68</td>
<td>75</td>
</tr>
<tr>
<td>First Quartile</td>
<td>68</td>
<td>65</td>
<td>71.25</td>
<td>77</td>
</tr>
<tr>
<td>Median</td>
<td>71</td>
<td>66</td>
<td>73.5</td>
<td>78.5</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>73</td>
<td>67.75</td>
<td>74.75</td>
<td>79.75</td>
</tr>
<tr>
<td>Maximum</td>
<td>75</td>
<td>69</td>
<td>77</td>
<td>81</td>
</tr>
<tr>
<td>IQR</td>
<td>5</td>
<td>2.75</td>
<td>3.5</td>
<td>2.75</td>
</tr>
<tr>
<td>1.5(IQR)</td>
<td>7.5</td>
<td>4.125</td>
<td>5.25</td>
<td>4.125</td>
</tr>
<tr>
<td>Lower Limit</td>
<td>60.5</td>
<td>60.875</td>
<td>66</td>
<td>72.875</td>
</tr>
<tr>
<td>Upper Limit</td>
<td>80.5</td>
<td>71.875</td>
<td>80</td>
<td>83.875</td>
</tr>
</tbody>
</table>

There are no outliers for any of the cell-phone services.

e. A box plot created using Excel’s Box and Whisker Statistical Chart follows.

The box plots show that Verizon is the best cell-phone service provider in terms of overall customer satisfaction. Verizon’s lowest rating is better than the highest AT&T and Sprint ratings and is better than 75% of the T-Mobile ratings. Sprint shows the lowest customer satisfaction ratings among the four services.

54. Excel’s AVERAGE, MIN, QUARTILE.EXC, and MAX functions provided the following results; values for IQR and the upper and lower limits are also shown.

**Personal Vehicles (1000s)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>173.24</td>
</tr>
<tr>
<td>Minimum</td>
<td>21</td>
</tr>
</tbody>
</table>
Descriptive Statistics: Numerical Measures

First Quartile 38.5  
Second Quartile 89.5  
Third Quartile 232  
Maximum 995  
IQR 193.5  
1.5(IQR) 290.25  
Lower Limit -251.75  
Upper Limit 522.25

a. Mean = 173.24 and median (second quartile) = 89.5  
b. First quartile = 38.5 and the third quartile = 232  
c. 21, 38.5, 89.5, 232, 995  
d. A box plot created using Excel’s Box and Whisker Statistical Chart follows.

The box plot shows the distribution of number of personal vehicle crossings is skewed to the right (positive). Three ports of entry are considered outliers:

NY: Buffalo-Niagara Falls 707
TX: El Paso 807
CA: San Ysidro 995
55. a.

\[
\begin{align*}
\sum x_i &= 40 & \bar{x} &= \frac{40}{5} = 8 \\
\sum y_i &= 230 & \bar{y} &= \frac{230}{5} = 46 \\
\sum (x_i - \bar{x})(y_i - \bar{y}) &= -240 & \sum (x_i - \bar{x})^2 &= 118 & \sum (y_i - \bar{y})^2 &= 520 \\
s_{xy} &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-240}{5-1} = -60 \\
s_x &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{118}{5-1}} = 5.4314 \\
s_y &= \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{520}{5-1}} = 11.4018 \\
r_{xy} &= \frac{s_{xy}}{s_x s_y} = \frac{-60}{(5.4314)(11.4018)} = -.969
\end{align*}
\]

Sample covariance = -60
The negative value of the sample covariance indicates a negative linear relationship.
Sample correlation coefficient = -.969
There is a strong negative linear relationship.

b. Negative relationship
56. a.

b. Positive relationship

c/d. \( \Sigma x_i = 80 \) \( \bar{x} = \frac{80}{5} = 16 \) \( \Sigma y_i = 50 \) \( \bar{y} = \frac{50}{5} = 10 \)

\[ \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 106 \quad \Sigma (x_i - \bar{x})^2 = 272 \quad \Sigma (y_i - \bar{y})^2 = 86 \]

\[ s_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{106}{5 - 1} = 26.5 \]

Sample covariance = 26.5

The positive value of the sample covariance indicates a positive linear relationship.

\[ s_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{272}{5 - 1}} = 8.2462 \]

\[ s_y = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n - 1}} = \sqrt{\frac{86}{5 - 1}} = 4.6368 \]

\[ r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{26.5}{(8.2462)(4.6368)} = .693 \]

Sample correlation coefficient = .693 which indicates a moderately strong positive linear relationship
58. Let $x = \text{miles per hour}$ and $y = \text{miles per gallon}$

\[
\begin{align*}
\Sigma x_i &= 420 \quad \bar{x} = \frac{420}{10} = 42 \\
\Sigma y_i &= 270 \quad \bar{y} = \frac{270}{10} = 27 \\
\Sigma (x_i - \bar{x})(y_i - \bar{y}) &= -475 \\
\Sigma (x_i - \bar{x})^2 &= 1660 \\
\Sigma (y_i - \bar{y})^2 &= 164
\end{align*}
\]

\[
s_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-475}{10-1} = -52.778
\]

\[
s_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1660}{10-1}} = 13.581
\]

\[
s_y = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{164}{10-1}} = 4.268
\]

\[
r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-52.778}{(13.581)(4.268)} = -.91
\]

A strong negative linear relationship exists. For driving speeds between 25 and 60 miles per hour, higher speeds are associated with lower miles per gallon.

60. a. % Return of DJIA versus Russell 1000

\[
\text{DJIA: } \bar{x} = \frac{\sum x_i}{n} = \frac{227.57}{25} = 9.10 \\
s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{5672.61}{24}} = 15.37
\]

\[
\text{Russell 1000: } \bar{x} = \frac{\sum x_i}{n} = \frac{227.29}{25} = 9.09 \\
s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{7679.81}{24}} = 17.89
\]

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
c. \[ r_{xy} = \frac{s_{xy}}{S_xS_y} = \frac{263.611}{(15.37)(17.89)} = .959 \]

d. Based on this sample, the two indexes are very similar. They have a strong positive correlation. The variance of the Russell 1000 is slightly larger than that of the DJIA.

62. The data in ascending order follow.

<table>
<thead>
<tr>
<th>Position</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
</tr>
</tbody>
</table>

a. The mean is 2.95 and the median is 3.

b. \[ L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (20 + 1) = 5.25 \]

First quartile or 25th percentile = \[ 1 + .25(1 - 1) = 1 \]

\[ L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (20 + 1) = 15.75 \]

Third quartile or 75th percentile = \[ 4 + .75(5 - 4) = 4.75 \]

c. The range is 7 and the interquartile range is 4.75 – 1 = 3.75.

d. The variance is 4.37 and standard deviation is 2.09.

e. Because most people dine out a relatively few times per week and a few families dine out very frequently, we would expect the data to be positively skewed. The skewness measure of 0.34 indicates the data are somewhat skewed to the right.

f. The lower limit is –4.625 and the upper limit is 10.375. No values in the data are less than the lower limit or greater than the upper limit, so there are no outliers.

64. a. The mean and median patient wait times for offices with a wait tracking system are 17.2 and 13.5, respectively. The mean and median patient wait times for offices without a wait tracking system are 29.1 and 23.5, respectively.

b. The variance and standard deviation of patient wait times for offices with a wait tracking system are 86.2 and 9.3, respectively. The variance and standard deviation of patient wait times for offices without a wait tracking system are 275.7 and 16.6, respectively.

c. Offices with a wait tracking system have substantially shorter patient wait times than offices without a wait tracking system.
d. \( z = \frac{37 - 29.1}{16.6} = 0.48 \)

e. \( z = \frac{37 - 17.2}{9.3} = 2.13 \)

As indicated by the positive z–scores, both patients had wait times that exceeded the means of their respective samples. Even though the patients had the same wait time, the z–score for the sixth patient in the sample who visited an office with a wait tracking system is much larger because that patient is part of a sample with a smaller mean and a smaller standard deviation.

f. The z–scores for all patients follow.

<table>
<thead>
<tr>
<th>Without Wait Tracking System</th>
<th>With Wait Tracking System</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.31</td>
<td>1.49</td>
</tr>
<tr>
<td>2.28</td>
<td>-0.67</td>
</tr>
<tr>
<td>-0.73</td>
<td>-0.34</td>
</tr>
<tr>
<td>-0.55</td>
<td>0.09</td>
</tr>
<tr>
<td>0.11</td>
<td>-0.56</td>
</tr>
<tr>
<td>0.90</td>
<td>2.13</td>
</tr>
<tr>
<td>-1.03</td>
<td>-0.88</td>
</tr>
<tr>
<td>-0.37</td>
<td>-0.45</td>
</tr>
<tr>
<td>-0.79</td>
<td>-0.56</td>
</tr>
<tr>
<td>0.48</td>
<td>-0.24</td>
</tr>
</tbody>
</table>

The z–scores do not indicate the existence of any outliers in either sample.

66. a. \( \bar{x} = \frac{\sum x_i}{n} = \frac{20665}{50} = 413.3 \) This is slightly higher than the mean for the study.

b. \( s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{69424.5}{49}} = 37.64 \)

c. \( L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (51) = 12.75 \) First Quartile or \( Q_1 = 374 + .75(384-374) = 381.5 \)

\( L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (51) = 38.25 \) Third Quartile or \( Q_3 = 445 + .25(445-445) = 445 \)

IQR = 445 – 381.5 = 63.5

LL = \( Q_1 - 1.5 \) IQR = 381.5 – 1.5(63.5) = 286.25

UL = \( Q_3 + 1.5 \) IQR = 445 + 1.5(63.5) = 540.25

There are no outliers.
68. Excel’s MIN, QUARTILE.EXC, and MAX functions provided the following results; values for the IQR and the upper and lower limits are also shown.

<table>
<thead>
<tr>
<th>Annual Household Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>First Quartile</td>
</tr>
<tr>
<td>Second Quartile</td>
</tr>
<tr>
<td>Third Quartile</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>IQR</td>
</tr>
<tr>
<td>1.5(IQR)</td>
</tr>
<tr>
<td>Lower Limit</td>
</tr>
<tr>
<td>Upper Limit</td>
</tr>
</tbody>
</table>

a. The data in ascending order follow:

\[
\begin{align*}
46.5 & \quad 48.7 & \quad 49.4 & \quad 51.2 & \quad 51.3 & \quad 51.6 & \quad 52.1 & \quad 52.1 & \quad 52.2 & \quad 52.4 & \quad 52.5 & \quad 52.9 & \quad 53.4 & \quad 64.5 \\
\end{align*}
\]

\[
\begin{align*}
L_{50} = \frac{p}{100} (n + 1) = \frac{50}{100} (14 + 1) = 7.5
\end{align*}
\]

Median or 50th percentile = 52.1 + .5(52.1 – 52.1) = 52.1

b. Percentage change = \[
\frac{52.1 - 55.5}{55.5} \times 100 = -6.1\%
\]

c. \[
\begin{align*}
L_{25} = \frac{p}{100} (n + 1) = \frac{25}{100} (14 + 1) = 3.75
\end{align*}
\]

25th percentile = 49.4 + .75(51.2 – 49.4) = 50.75

\[
\begin{align*}
L_{75} = \frac{p}{100} (n + 1) = \frac{75}{100} (14 + 1) = 11.25
\end{align*}
\]

75th percentile = 52.5 + .25(52.9 – 52.5) = 52.6

d. 46.5 50.75 52.1 52.6 64.5

e. \[
\begin{align*}
\bar{x} = \frac{\sum{x_i}}{n} = \frac{730.8}{14} = 52.2
\end{align*}
\]

\[
\begin{align*}
s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{208.12}{13} = 16.0092
\end{align*}
\]

\[
\begin{align*}
s = \sqrt{16.0092} = 4.0012
\end{align*}
\]

The z-scores = \[
\frac{x_i - \bar{x}}{s}
\]

are shown below:
The last household income (64.5) has a z-score > 3 and is an outlier.

Lower Limit = \( Q_1 - 1.5(\text{IQR}) = 50.75 - 1.5(52.6 - 50.75) = 47.98 \)

Upper Limit = \( Q_3 + 1.5(\text{IQR}) = 52.6 + 1.5(52.6 - 50.75) = 55.38 \)

Using this approach the first observation (46.5) and the last observation (64.5) would be considered outliers.

The two approaches will not always provide the same results.

70. a. \( \bar{x} = \frac{\Sigma x_i}{n} = \frac{4368}{12} = 364 \) rooms

b. \( \bar{y} = \frac{\Sigma y_i}{n} = \frac{5484}{12} = \$457 \)

c.

It is difficult to see much of a relationship. When the number of rooms becomes larger, there is no indication that the cost per night increases. The cost per night may even decrease slightly.

\[
\begin{array}{cccccccc}
 x_i & y_i & (x_i - \bar{x}) & (y_i - \bar{y}) & (x_i - \bar{x})^2 & (y_i - \bar{y})^2 & (x_i - \bar{x})(y_i - \bar{y}) \\
 220 & 499 & -144 & 42 & 20,736 & 1,764 & -6,048 \\
 727 & 340 & 363 & -117 & 131,769 & 13,689 & -42,471 \\
 285 & 585 & -79 & 128 & 6,241 & 16,384 & -10,112 \\
 273 & 495 & -91 & 38 & 8,281 & 1,444 & -3,458 \\
 145 & 495 & -219 & 38 & 47,961 & 1,444 & -8,322 \\
 213 & 279 & -151 & -178 & 22,801 & 31,684 & 26,878 \\
 398 & 279 & -34 & -178 & 1,156 & 31,684 & -6,052 \\
 343 & 455 & -21 & -2 & 441 & 4 & 42 \\
 250 & 595 & -114 & 138 & 12,996 & 19,044 & -15,732 \\
 414 & 367 & 50 & -90 & 2,500 & 8,100 & -4,500 \\
\end{array}
\]
There is evidence of a slightly negative linear association between the number of rooms and the cost per night for a double room. Although this is not a strong relationship, it suggests that the higher room rates tend to be associated with the smaller hotels.

This tends to make sense when you think about the economies of scale for the larger hotels. Many of the amenities in terms of pools, equipment, spas, restaurants, and so on exist for all hotels in the Travel + Leisure top 50 hotels in the world. The smaller hotels tend to charge more for the rooms. The larger hotels can spread their fixed costs over many room and may actually be able to charge less per night and still achieve and nice profit. The larger hotels may also charge slightly less in an effort to obtain a higher occupancy rate. In any case, it appears that there is a slightly negative linear association between the number of rooms and the cost per night for a double room at the top hotels.

### 72. a.

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$y_i$</th>
<th>$(x_i - \bar{x})$</th>
<th>$(y_i - \bar{y})$</th>
<th>$(x_i - \bar{x})^2$</th>
<th>$(y_i - \bar{y})^2$</th>
<th>$(x_i - \bar{x})(y_i - \bar{y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.407</td>
<td>.422</td>
<td>-.1458</td>
<td>-0.0881</td>
<td>.0213</td>
<td>.0078</td>
<td>.0128</td>
</tr>
<tr>
<td>.429</td>
<td>.586</td>
<td>-.1238</td>
<td>.0759</td>
<td>.0153</td>
<td>.0058</td>
<td>-.0094</td>
</tr>
<tr>
<td>.417</td>
<td>.546</td>
<td>-.1358</td>
<td>.0359</td>
<td>.0184</td>
<td>.0013</td>
<td>-.0049</td>
</tr>
<tr>
<td>.569</td>
<td>.500</td>
<td>.0162</td>
<td>-.0101</td>
<td>.0003</td>
<td>.0001</td>
<td>-.0002</td>
</tr>
<tr>
<td>.569</td>
<td>.457</td>
<td>.0162</td>
<td>-.0531</td>
<td>.0003</td>
<td>.0028</td>
<td>-.0009</td>
</tr>
<tr>
<td>.533</td>
<td>.463</td>
<td>-.0198</td>
<td>-.0471</td>
<td>.0004</td>
<td>.0022</td>
<td>.0009</td>
</tr>
<tr>
<td>.724</td>
<td>.617</td>
<td>.1712</td>
<td>.1069</td>
<td>.0293</td>
<td>.0114</td>
<td>.0183</td>
</tr>
<tr>
<td>.500</td>
<td>.540</td>
<td>-.0528</td>
<td>.0299</td>
<td>.0028</td>
<td>.0009</td>
<td>-.0016</td>
</tr>
<tr>
<td>.577</td>
<td>.549</td>
<td>.0242</td>
<td>.0389</td>
<td>.0006</td>
<td>.0015</td>
<td>.0009</td>
</tr>
</tbody>
</table>
b. There is a low positive correlation between a major league baseball team’s winning percentage
during spring training and its winning percentage during the regular season. The spring training
record should not be expected to be a good indicator of how a team will play during the regular
season.

Spring training consists of practice games between teams with the outcome as to who wins or who
loses not counting in the regular season standings or affecting the chances of making the playoffs.
Teams use spring training to help players regain their timing and evaluate new players. Substitutions
are frequent with the regular or better players rarely playing an entire spring training game. Winning
is not the primary goal in spring training games. A low correlation between spring training winning
percentage and regular season winning percentage should be anticipated.

### 74.

<table>
<thead>
<tr>
<th>$w_i$</th>
<th>$x_i$</th>
<th>$w_i x_i$</th>
<th>$x_i - \bar{x}$</th>
<th>$(x_i - \bar{x})^2$</th>
<th>$w_i (x_i - \bar{x})^2$</th>
<th>$(x_i - \bar{x})^2$</th>
<th>$w_i (x_i - \bar{x})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>47</td>
<td>470</td>
<td>-13.68</td>
<td>187.1424</td>
<td>1871.42</td>
<td>187.26</td>
<td>1872.58</td>
</tr>
<tr>
<td>40</td>
<td>52</td>
<td>2080</td>
<td>-8.68</td>
<td>75.3424</td>
<td>3013.70</td>
<td>75.42</td>
<td>3016.62</td>
</tr>
<tr>
<td>150</td>
<td>57</td>
<td>8550</td>
<td>-3.68</td>
<td>13.5424</td>
<td>2031.36</td>
<td>13.57</td>
<td>2036.01</td>
</tr>
<tr>
<td>175</td>
<td>62</td>
<td>10850</td>
<td>+1.32</td>
<td>1.7424</td>
<td>304.92</td>
<td>1.73</td>
<td>302.98</td>
</tr>
<tr>
<td>75</td>
<td>67</td>
<td>5025</td>
<td>+6.32</td>
<td>39.9424</td>
<td>2995.68</td>
<td>39.89</td>
<td>2991.69</td>
</tr>
<tr>
<td>15</td>
<td>72</td>
<td>1080</td>
<td>+11.32</td>
<td>128.1424</td>
<td>1922.14</td>
<td>128.05</td>
<td>1920.71</td>
</tr>
<tr>
<td>10</td>
<td>77</td>
<td>770</td>
<td>+16.32</td>
<td>266.3424</td>
<td>2663.42</td>
<td>266.20</td>
<td>2662.05</td>
</tr>
<tr>
<td>475</td>
<td>28.825</td>
<td>14,802.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Columns 5 and 6 are calculated with rounding, while columns 7 and 8 are based on unrounded calculations.

a. $\bar{x} = \frac{28,825}{475} = 60.68$

b. $s^2 = \frac{14,802.64}{474} = 31.23$

$s = \sqrt{31.23} = 5.59$
Chapter 4
Introduction to Probability

Solutions:

2. \[ \binom{6}{3} = \frac{6!}{3!3!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(3 \cdot 2 \cdot 1)} = 20 \]

ABC  ACE  BCD  BEF
ABD  ACF  BCE  CDE
ABE  ADE  BCF  CDF
ABF  ADF  BDE  CEF
ACD  AEF  BDF  DEF

4. a. 

<table>
<thead>
<tr>
<th>1st Toss</th>
<th>2nd Toss</th>
<th>3rd Toss</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H</td>
<td>(H,H,H)</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>(H,H,T)</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>(H,T,H)</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>(H,T,T)</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>(T,H,H)</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>(T,H,T)</td>
</tr>
<tr>
<td>H</td>
<td>T</td>
<td>(T,T,H)</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>(T,T,T)</td>
</tr>
</tbody>
</table>

b. Let: H be head and T be tail

(H,H,H) (T,H,H)
(H,H,T) (T,H,T)
(H,T,H) (T,T,H)
(H,T,T) (T,T,T)

c. The outcomes are equally likely, so the probability of each outcome is 1/8.

6. \[ P(E_1) = .40, \quad P(E_2) = .26, \quad P(E_3) = .34 \]

The relative frequency method was used.

8. a. There are four outcomes possible for this 2-step experiment; planning commission positive - council approves; planning commission positive - council disapproves; planning commission negative - council approves; planning commission negative - council disapproves.

b. Let p = positive, n = negative, a = approves, and d = disapproves
9. \[
\binom{50}{4} = \frac{50!}{4!46!} = \frac{50 \cdot 49 \cdot 48 \cdot 47}{4 \cdot 3 \cdot 2 \cdot 1} = 230,300
\]

10. a. Using the table provided, 86.5% of Delta flights arrive on time.

\[P(\text{on-time arrival}) = .865\]

b. Three of the 10 airlines have less than two mishandled baggage reports per 1000 passengers.

\[P(\text{Less than 2}) = \frac{3}{10} = .30\]

c. Five of the 10 airlines have more than one customer complaints per 1000 passengers.

\[P(\text{more than 1}) = \frac{5}{10} = .50\]

d. \[P(\text{not on time}) = 1 - P(\text{on time}) = 1 - .871 = .129\]

12. a. Step 1: Use the counting rule for combinations:

\[
\binom{59}{5} = \frac{59!}{5!(59-5)!} = \frac{(59\times58\times57\times56\times55)}{(5\times4\times3\times2\times1)} = 5,006,386
\]

Step 2: There are 35 ways to select the red Powerball from digits 1 to 35.

Total number of Powerball lottery outcomes: \(5,006,386 \times 35 = 175,223,510\)

b. Probability of winning the lottery: 1 chance in 175,223,510
14. a. \( P(E_2) = \frac{1}{4} \)

b. \( P(\text{any 2 outcomes}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \)

c. \( P(\text{any 3 outcomes}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \)

15. a. \( S = \{\text{ace of clubs, ace of diamonds, ace of hearts, ace of spades}\} \)

b. \( S = \{2 \text{ of clubs, 3 of clubs, \ldots, 10 of clubs, J of clubs, Q of clubs, K of clubs, A of clubs}\} \)

c. There are 12; jack, queen, or king in each of the four suits.

d. For a: \( \frac{4}{52} = \frac{1}{13} = .08 \)

For b: \( \frac{13}{52} = \frac{1}{4} = .25 \)

For c: \( \frac{12}{52} = .23 \)

16. a. \( (6)(6) = 36 \) sample points

b. 

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 \\
3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
4 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 6 & 7 & 8 & 9 & 10 & 11 \\
6 & 7 & 8 & 9 & 10 & 11 & 12 \\
\end{array}
\]

c. \( \frac{6}{36} = \frac{1}{6} \)

d. \( \frac{10}{36} = \frac{5}{18} \)

e. No. \( P(\text{odd}) = \frac{18}{36} = P(\text{even}) = \frac{18}{36} \) or 1/2 for both.

f. Classical. A probability of 1/36 is assigned to each experimental outcome.
17. a. (4,6), (4,7), (4,8)
   b. .05 + .10 + .15 = .30
   c. (2,8), (3,8), (4,8)
   d. .05 + .05 + .15 = .25
   e. Both are over budget only with (4,8) and therefore probability = .15

18. a. Let \( C \) = corporate headquarters located in California

\[ P(C) = \frac{53}{500} = .106 \]

b. Let \( N \) = corporate headquarters located in New York
   \( T \) = corporate headquarters located in Texas

\[ P(N) = \frac{50}{500} = .100 \]
\[ P(T) = \frac{52}{500} = .104 \]

Located in California, New York, or Texas

\[ P(C) + P(N) + P(T) = .106 + .100 + .104 = .31 \]

c. Let \( A \) = corporate headquarters located in one of the eight states

Total number of companies with corporate headquarters in the eight states = 283

\[ P(A) = \frac{283}{500} = .566 \]

Over half the Fortune 500 companies have corporate headquartered located in these eight states.

20. a.

<table>
<thead>
<tr>
<th>Age</th>
<th>Experimental Outcome</th>
<th>Financially Independent</th>
<th>Number of Responses</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>E_1</td>
<td>16 to 20</td>
<td>191</td>
<td>191/944 = 0.2023</td>
<td></td>
</tr>
<tr>
<td>E_2</td>
<td>21 to 24</td>
<td>467</td>
<td>467/944 = 0.4947</td>
<td></td>
</tr>
<tr>
<td>E_3</td>
<td>25 to 27</td>
<td>244</td>
<td>244/944 = 0.2585</td>
<td></td>
</tr>
<tr>
<td>E_4</td>
<td>28 or older</td>
<td>42</td>
<td>42/944 = 0.0445</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>944</td>
<td></td>
</tr>
</tbody>
</table>

b. \( P(\text{Age <25}) = P(E_1) + P(E_2) = .2023 + .4947 = .6970 \)

c. \( P(\text{Age >24}) = P(E_3) + P(E_4) = .2585 + .0445 = .3030 \)

d. The probability of being financially independent before the age of 25, .6970, seems high given the general economic conditions. It appears that the teenagers who responded to this survey may have unrealistic expectations about becoming financially independent at a relatively young age.
22. \(E_1 = E_2 = E_3 = E_4 = E_5 = 0.2\) for each
   A = \(E_1, E_2\)
   B = \(E_3, E_4\)
   C = \(E_2, E_3, E_5\)

   a. \(P(A) = .2+.2=.40, P(B) = .2+.2=.40, P(C) = .2+.2+.2=.60\)

   b. \(P(A \cup B) = P(E_1, E_2, E_3, E_4) = .80. Yes P(A \cup B) = P(A) + P(B) because they do not have any outcomes in common.\)

   c. \(A^c = \{E_3, E_4, E_5\} \quad C^c = \{E_1, E_4\} \quad P(A^c) = .60 \quad P(C^c) = .40\)

   d. \(A \cup B^c = \{E_1, E_2, E_5\} \quad P(A \cup B^c) = .60\)

   e. \(P(B \cup C) = P(E_2, E_3, E_4, E_5) = .80\)

23. a. \(P(A) = P(E_1) + P(E_4) + P(E_6) = .05 + .25 + .10 = .40\)

   \(P(B) = P(E_2) + P(E_4) + P(E_7) = .20 + .25 + .05 = .50\)

   \(P(C) = P(E_2) + P(E_3) + P(E_5) + P(E_7) = .20 + .20 + .15 + .05 = .60\)

   b. \(A \cup B = \{E_1, E_2, E_4, E_6, E_7\}\)

   \(P(A \cup B) = P(E_1) + P(E_2) + P(E_4) + P(E_6) + P(E_7) = .05 + .20 + .15 + .10 = .50\)

   OR \(P(A \cup B) = P(A) + P(B) - P(A \cap B)\)

   c. \(A \cap B = \{E_4\}\)

   \(P(A \cap B) = P(E_4) = .25\)

   d. Yes, they are mutually exclusive because they do not have any outcomes in common (there is no intersection between them).

   e. \(B^c = \{E_1, E_3, E_5, E_6\}; \quad P(B^c) = P(E_1) + P(E_3) + P(E_5) + P(E_6) = .05 + .20 + .15 + .10 = .50\)

24. Let \(E = \) experience exceeded expectations
    \(M = \) experience met expectations
    \(S = \) experience fell short of expectations
    \(N = \) no response

   a. Percentage of respondents that said their experience exceeded expectations
      \[= 1 - \text{all other responses} = 1 - \{ P(N) + P(S) + P(M) \} = 1 - (.04+.26+.65) = .05 \]

   \(P(E) = .05\)

   b. \(P(M \cup E) = P(M) + P(E) = .65 + .05 = .70\)

26. a. Let \(D = \) Domestic Equity Fund

   \(P(D) = 16/25 = .64\)

   b. Let \(A = 4-\) or 5-star rating

   13 funds were rated 3-star of less; thus, \(25 - 13 = 12\) funds must be 4-star or 5-star.

   \(P(A) = 12/25 = .48\)
Chapter 4

c. 7 Domestic Equity funds were rated 4-star and 2 were rated 5-star. Thus, 9 funds were Domestic Equity funds and were rated 4-star or 5-star

\[ P(D \cap A) = \frac{9}{25} = .36 \]

d. \[ P(D \cup A) = P(D) + P(A) - P(D \cap A) \]

\[ = .64 + .48 - .36 = .76 \]

28. Let: \( B = \text{rented a car for business reasons} \)
\( P = \text{rented a car for personal reasons} \)

a. \[ P(B \cup P) = P(B) + P(P) - P(B \cap P) \]

\[ = .54 + .458 - .30 = .698 \]

b. \( P(\text{Neither}) = 1 - .698 = .302 \)

30. a. \[ P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{.40}{.60} = .6667 \]

b. \[ P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{.40}{.50} = .80 \]

c. No because \( P(A | B) \neq P(A) \)

32. a. Dividing each entry in the table by 500 yields the following (rounding to two digits):

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.210</td>
<td>0.282</td>
<td>0.492</td>
</tr>
<tr>
<td>Women</td>
<td>0.186</td>
<td>0.322</td>
<td>0.508</td>
</tr>
<tr>
<td>Totals</td>
<td>0.396</td>
<td>0.604</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Let \( M = \text{18-34-year-old man} \), \( W = \text{18-34-year-old woman} \), \( Y = \text{responded yes} \), \( N = \text{responded no} \)

b. \( P(M) = .492, P(W) = .508 \)
\( P(Y) = .396, P(N) = .604 \)

c. \( P(Y | M) = .210/ .492 = .4268 \)

d. \( P(Y | W) = .186/ .508 = .3661 \)

e. \( P(Y) = .396/1 = .396 \)

f. \( P(M) = .492 \) in the sample. Yes, this seems like a good representative sample based on gender.
33. a. 

<table>
<thead>
<tr>
<th>Intended Enrollment Status</th>
<th>Full-Time</th>
<th>Engineering</th>
<th>Other</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-Time</td>
<td>.2697</td>
<td>.1510</td>
<td>.1923</td>
<td>.6130</td>
</tr>
<tr>
<td>Part-Time</td>
<td>.1149</td>
<td>.1234</td>
<td>.1487</td>
<td>.3870</td>
</tr>
<tr>
<td>Totals</td>
<td>.3847</td>
<td>.2743</td>
<td>.3410</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

b. Let $B =$ undergraduate major in business 
   $E =$ undergraduate major in engineering 
   $O =$ other undergraduate major 
   $F =$ full-time enrollment

$P(B) = .3847$, $P(E) = .2743$, and $P(O) = .3410$, so business is the undergraduate major that produces the most potential MBA students.

c. $P(F | B) = \frac{P(F \cap B)}{P(B)} = \frac{.1510}{.3847} = .3912$

d. $P(F | B) = \frac{P(F \cap B)}{P(B)} = \frac{.2697}{.3847} = .7012$

e. For Independent, $P(F/B) = P(F)$
   $P(F/B) = .7012$
   $P(F) = .6130$

Since, $P(F/B) \neq P(F)$, the events are not independent

34. a. Let $O =$ flight arrives on time 
   $L =$ flight arrives late 
   $J =$ Jet Blue flight 
   $N =$ United flight 
   $U =$ US Airways flight

Given: 76.8% of Jetblue arrives on time $= P(O \mid J) = .768$
   71.5% of United arrives on time $= P(O \mid N) = .715$
   82.2% of USAirways arrives on time $= P(O \mid U) = .822$
   $P(J) = .30$ 
   $P(N) = .32$ 
   $P(U) = .38$

Joint probabilities using the multiplication law

$P(J \cap O) = P(J)P(O \mid J) = (.30)(.768) = .2304$
$P(N \cap O) = P(N)P(O \mid N) = (.32)(.715) = .2288$
$P(U \cap O) = P(U)P(O \mid U) = (.38)(.822) = .3124$

With the marginal probabilities $P(J) = .30$, $P(N) = .32$, and $P(U) = .38$ given, the joint probability table can then be shown as follows.

<table>
<thead>
<tr>
<th></th>
<th>On time</th>
<th>Late</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jet Blue</td>
<td>.2304</td>
<td>.0696</td>
<td>.30</td>
</tr>
<tr>
<td>United</td>
<td>.2288</td>
<td>.0912</td>
<td>.32</td>
</tr>
</tbody>
</table>
Chapter 4

<table>
<thead>
<tr>
<th>US Airways</th>
<th>.31236</th>
<th>.06764</th>
<th>.38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>.77156</td>
<td>.22844</td>
<td>1.00</td>
</tr>
</tbody>
</table>

b. Using the joint probability table, the probability of an on-time flight is the marginal probability

\[ P(O) = .2304 + .2288 + .31236 = .77156 \]

c. Since US Airways has the highest percentage of flights into terminal C, US Airways with \( P(U) = .38 \) is the most likely airline for Flight 1382.

d. From the joint probability table, \( P(L) = .22844 \)

\[ P(J \mid L) = \frac{P(J \cap L)}{P(L)} = \frac{.0696}{.2284} = .3047 \]

\[ P(N \mid L) = \frac{P(N \cap L)}{P(L)} = \frac{.0912}{.2284} = .3992 \]

\[ P(U \mid L) = \frac{P(U \cap L)}{P(L)} = \frac{.0676}{.2284} = .2961 \]

Most likely airline for Flight 1382 is now United with a probability of .3992. US Airways is now the least likely airline for this flight with a probability of .2961.

36. a. Let \( A \) = makes 1st free throw
\( B \) = makes 2nd free throw
Probability of both shots made = \( P(A \cap B) \)
Assuming independence, \( P(A \cap B) = P(A)P(B) = (.93)(.93) = .8649 \)

b. At least one shot = Probability of one shot or the other, or both = \( P(A \cup B) \)
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) = (.93) + (.93) - .8649 = .9951 \]

c. Miss both \( 1 - P(A \cup B) = 1 - .9951 = .0049 \)

d. For the Portland Trail Blazers’ center with \( P(A) = P(B) = .58 \)
\[ P(A \cap B) = P(A)P(B) = (.58)(.58) = .3346 \]
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) = (.58) + (.58) - .3346 = .8236 \]
Miss both \( 1 - P(A \cup B) = 1 - .8236 = .1764 \)

Intentionally fouling the Portland Trail Blazers’ center is a better strategy than intentionally fouling Jamal Crawford.

38. Let \( Y \) = has a college degree
\( N \) = does not have a college degree
\( D \) = a delinquent student loan

a. From the table, \( P(Y) = .42 \)
b. From the table, \( P(N) = .58 \)

c. \[
P(D|Y) = \frac{P(D \cap Y)}{P(Y)} = \frac{.16}{.2} = 0.810
\]
d. \[
P(D|N) = \frac{P(D \cap N)}{P(N)} = \frac{.34}{.58} = 0.5862
\]
e. Individuals who obtained a college degree have a .3810 probability of a delinquent student loan while individuals who dropped out without obtaining a college degree have a .5862 probability of a delinquent student loan. Not obtaining a college degree will lead to a greater probability of struggling to payback the student loan and will likely lead to financial problems in the future.

39. a. Yes, since \( P(A_1 \cap A_2) = 0 \)

b. \[
P(A_1 \cap B) = P(A_1)P(B | A_1) = .40(.20) = 0.08
\]
\[
P(A_2 \cap B) = P(A_2)P(B | A_2) = .60(.05) = 0.03
\]
c. \[
P(B) = P(A_1 \cap B) + P(A_2 \cap B) = .08 + .03 = .11
\]
d. \[
P(A_1 | B) = \frac{.08}{.11} = 0.7273
\]
\[
P(A_2 | B) = \frac{.03}{.11} = 0.2727
\]

40. a. \[
P(B \cap A_1) = P(A_1)P(B | A_1) = (.20)(.50) = 0.10
\]
\[
P(B \cap A_2) = P(A_2)P(B | A_2) = (.50)(.40) = 0.20
\]
\[
P(B \cap A_3) = P(A_3)P(B | A_3) = (.30)(.30) = 0.09
\]

b. \[
P(A_2 | B) = \frac{.20}{.10 + .20 + .09} = 0.51
\]

c. 

| Events | \( P(A_i) \) | \( P(B | A_i) \) | \( P(A_i \cap B) \) | \( P(A_i | B) \) |
|--------|-------------|---------------|-----------------|-----------|
| \( A_1 \) | .20         | .50           | .10             | .26       |
| \( A_2 \) | .50         | .40           | .20             | .51       |
| \( A_3 \) | .30         | .30           | .09             | .23       |

42. \( M = \text{missed payment} \)
\( D_1 = \text{customer defaults} \)
\( D_2 = \text{customer does not default} \)

\[
P(D_1) = .05 \quad P(D_2) = .95 \quad P(M | D_2) = .2 \quad P(M | D_1) = 1
\]

a. \[
P(D_1 | M) = \frac{P(D_1)P(M | D_1)}{P(D_1)P(M | D_1) + P(D_2)P(M | D_2)} = \frac{.05(1)}{.05(1) + (.95)(.2)} = \frac{.05}{.24} = 0.21
\]
Alternative Solutions for Part a:

Bayes Table

| Events | $P(D_i)$ | $P(M | D_i)$ | $P(D_i \cap M)$ | $P(D_i | M)$ |
|--------|----------|--------------|-----------------|-------------|
| $D_1$  | .05      | 1.0          | .05             | .21         |
| $D_2$  | .95      | .20          | .19             | .79         |

From the Bayes Table, $P(D_1/M) = .21$

Probability Table

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$M'$</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>1*.05 = .05</td>
<td>0</td>
<td>.05</td>
</tr>
<tr>
<td>$D_2$</td>
<td>.2*.95 = .19</td>
<td>.76</td>
<td>.95</td>
</tr>
</tbody>
</table>

From the Probability Table, $P(S_1/B) = .05/2.4 = .21$

b. Yes, the probability of default is greater than .20.

44. $M = $ the current visitor to the ParFore website is a male
$F = $ the current visitor to the ParFore website is a female
$D = $ a visitor to the ParFore website previously visited the Dillard website

a. Using past history, $P(F) = .40$.

b. $P(M) = .60$, $P(D | F) = .30$, and $P(D | M) = .10$

$$P(F | D) = \frac{P(F)P(D | F)}{P(F)P(D | F) + P(M)P(D | M)} = \frac{(.40)(.30)}{(.40)(.30) + (.60)(.10)} = .6667$$

The revised probability that the current visitor is a female is .6667.
ParFore should display the special offer that appeals to female visitors.

Alternative Solutions for Part b:

Bayes Table

| Events | $P(\text{gender})$ | $P(D | \text{gender})$ | $P(\text{gender} \cap D)$ | $P(\text{gender} | D)$ |
|--------|---------------------|-------------------------|---------------------------|-----------------------|
| $F$    | .4                  | .3                      | .12                       | .6667                 |
| $M$    | .6                  | .1                      | .06                       | .3333                 |

From the Bayes Table, $P(F/D) = .6667$

Probability Table

<table>
<thead>
<tr>
<th></th>
<th>$D$</th>
<th>$D'$</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>.3* .4 = .12</td>
<td>.28</td>
<td>.4</td>
</tr>
<tr>
<td>$M$</td>
<td>.6* .06 = .36</td>
<td>.36</td>
<td>.6</td>
</tr>
</tbody>
</table>
From the Probability Table, \( P(F/D) = \frac{.12}{.18} = .6667 \)

46. a. \( 422 + 181 + 80 + 121 + 201 = 1005 \) respondents

b. Most frequent response a day or less; \( P = \frac{422}{1005} = .4199 \)

c. \( 201/1005 = .20 \)

d. Responses of 2 days, 3 days, and 4 or more days = \( 181 + 80 + 121 = 382 \)

\( P = \frac{382}{1005} = .3801 \)

48. a. There are a total of 1364 responses. Dividing each entry by 1364 provides the following joint probability table.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>.2896</td>
<td>.2133</td>
<td>.5029</td>
</tr>
<tr>
<td>Male</td>
<td>.2368</td>
<td>.2603</td>
<td>.4971</td>
</tr>
<tr>
<td>Total</td>
<td>.5264</td>
<td>.4736</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

b. The marginal probability of a female is .5029 from above.

c. Let \( A = \) uses social media and other websites to voice opinions about television programs \( F = \) female respondent

\[
P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{.2896}{.5029} = .5758 \quad \text{OR} \quad \text{From raw data,} \quad \frac{395}{686} = .5758
\]

d. For Independent, \( P(A/F) = P(A) \)

\[
P(A/F) = .5758 \quad P(A) = .5264
\]

Since \( P(A/F) \neq P(A) \), the events are not independent

50. a. Probability of the event \( = P(\text{average}) + P(\text{above average}) + P(\text{excellent}) \)

\[
= \frac{11}{50} + \frac{14}{50} + \frac{13}{50} = \frac{22}{50} + \frac{26}{50} = .76
\]

b. Probability of the event \( = P(\text{poor}) + P(\text{below average}) \)

\[
= \frac{4}{50} + \frac{8}{50} = .24
\]

52. a.
Although the columns appear to add up to .4005 and .5995, the actual calculation resulting from using raw column totals results in \(808/2018 = .4004\) and \(1210/2018 = .5996\). Students may have either answer, and both could be considered correct.

b. Marginal probability .2022

c. \(.2245 + .1283 + .1090 = .4618\)

d. Marginal probability \(= 808/2018 = .4004\) See note above.

54. a. \(P(\text{Not Okay}) = .1485 + .2273 + .4008 = .7766\)

b. \(P(30 – 49) = .2273 + .0907 = 3.180\)

\[
P(\text{Okay} | 30 – 49) = \frac{P(\text{Okay} \cap 30 – 49)}{P(30 – 49)} = \frac{0.0907}{3.180} = 0.2852
\]

c. \[P(50+ | \text{Not Okay}) = \frac{P(50+ \cap \text{Not Okay})}{P(\text{Not Okay})} = \frac{.4008}{.7766} = .5161
\]

d. The attitude about this practice is not independent of the age of the respondent. One way to show this follows.

\(P(\text{Okay}) = 1 – P(\text{Not Okay}) = 1 - .7766 = .2234\)

\[P(\text{Okay} | 30 – 49) = .2852
\]

Since \(P(\text{Okay} | 30 – 49) \neq P(\text{Okay})\), attitude is not independent of the age of the respondent.

e. \[P(\text{Not Okay} | 50+) = \frac{P(\text{Not Okay} \cap 50+)}{P(50+)} = \frac{.4008}{.4731} = .8472
\]

\[P(\text{Not Okay} | 18-29) = \frac{P(\text{Not Okay} \cap 18-29)}{P(18-29)} = \frac{.1485}{.2089} = .7109
\]

There is a higher probability the 50+ year olds will not be okay with this practice.

56. a. \(P(A) = 200/800 = .25\)

b. \(P(B) = 100/800 = .125\)

c. \(P(A \cap B) = 10/800 = .0125\)

d. \(P(A | B) = P(A \cap B)/P(B) = .0125/.125 = .10\) OR From raw data, \(10/100 = .10\)

e. No, \(P(A | B) \neq P(A)\) or \(.10 \neq .25\)
58. a. Let $A_1 =$ student studied abroad  
    $A_2 =$ student did not study abroad  
    $F =$ female student  
    $M =$ male student  

    $P(A_1) = .095$  
    $P(A_2) = 1 - P(A_1) = 1 - .095 = .905$  
    $P(F \mid A_1) = .60$  
    $P(F \mid A_2) = .49$  

    Tabular computations  

    \[
    \begin{array}{c|c|c|c|c}
    \text{Events} & P(A_i) & P(F \mid A_i) & P(A_i \cap F) & P(A_i \mid F) \\
    \hline
    A_1 & .095 & .60 & .0570 & .1139 \\
    A_2 & .905 & .49 & .4435 & .8861 \\
    \end{array}
    \]

    $P(F) = .5005$  
    $P(A_1 \mid F) = .1139$  

    b.  

    \[
    \begin{array}{c|c|c|c|c|c}
    \text{Events} & P(A_i) & P(M \mid A_i) & P(A_i \cap M) & P(A_i \mid M) \\
    \hline
    A_1 & .095 & .40 & .0380 & .0761 \\
    A_2 & .905 & .51 & .4615 & .9239 \\
    \end{array}
    \]

    $P(M) = .4995$  
    $P(A_1 \mid M) = .0761$  

    Alternative Solution for Part a and b:  

    Probability Table  

    \[
    \begin{array}{c|c|c|c|c|c}
    & F & M & \text{Totals} \\
    \hline
    A_1 & .6* .095 = & .4* .095 = & .095 & \text{.0570 or} \text{.0570} \\
    & .0570 & = .0380 & .095 \\
    A_2 & .49* .905 = & .51* .905 = & .905- .4435 or & .905 \text{.4435} \\
    & .4435 & .4615 & .905 \\
    \text{Totals} & 0.5005 & 0.4995 & 1.00 \\
    \end{array}
    \]

    From the Probability Table, $P(A_1 \mid F) = .0570/.5005 = .1139$ and $P(A_1 \mid M) = .0380/.4995 = .0761$  

    c. From above, $P(F) = .5005$ and $P(M) = .4995$, so almost 50/50 female and male full-time students.
Chapter 4

60. a.

\[
P(\text{ham} | \text{shipping}!) = \frac{P(\text{ham})P(\text{shipping}! | \text{ham})}{P(\text{ham})P(\text{shipping}! | \text{ham}) + P(\text{spam})P(\text{shipping}! | \text{spam})} \\
= \frac{(.90)(.0015)}{(.90)(.0015) + (.10)(.051)} = .2093
\]

If a message includes the word *shipping!*, the probability the message is spam is high (.7910), and so the message should be flagged as spam.

b.

\[
P(\text{spam} | \text{here}!) = \frac{P(\text{spam})P(\text{here}! | \text{spam})}{P(\text{spam})P(\text{here}! | \text{spam}) + P(\text{ham})P(\text{here}! | \text{ham})} \\
= \frac{\binom{10}{0.034}}{\binom{10}{0.034} + \binom{90}{0.022}} = .6320
\]

A message that includes the word *today!* is more likely to be spam. This is because \(P(\text{today}! | \text{spam})\) is larger than \(P(\text{here}! | \text{spam})\). Because *today!* occurs more often in unwanted messages (spam), it is easier to distinguish spam from ham in messages that include *today!*.

c.

\[
P(\text{spam} | \text{available}!) = \frac{P(\text{spam})P(\text{available}! | \text{spam})}{P(\text{spam})P(\text{available}! | \text{spam}) + P(\text{ham})P(\text{available}! | \text{ham})} \\
= \frac{\binom{10}{0.014}}{\binom{10}{0.014} + \binom{90}{0.041}} = .2750
\]

\[
P(\text{spam} | \text{fingertips}!) = \frac{P(\text{spam})P(\text{fingertips}! | \text{spam})}{P(\text{spam})P(\text{fingertips}! | \text{spam}) + P(\text{ham})P(\text{fingertips}! | \text{ham})} \\
= \frac{\binom{16}{0.014}}{\binom{16}{0.014} + \binom{90}{0.011}} = .5858
\]

A message that includes the word *fingertips!* is more likely to be spam.

d. It is easier to distinguish spam from ham when a word occurs more often in unwanted messages (spam) and/or less often in legitimate messages (ham).
Chapter 5
Discrete Probability Distributions

Solutions:

1. a. Head, Head (H,H)
   Head, Tail (H,T)
   Tail, Head (T,H)
   Tail, Tail (T,T)

   b. \( x \) = number of heads on two coin tosses

c. 

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Values of ( x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H,H)</td>
<td>2</td>
</tr>
<tr>
<td>(H,T)</td>
<td>1</td>
</tr>
<tr>
<td>(T,H)</td>
<td>1</td>
</tr>
<tr>
<td>(T,T)</td>
<td>0</td>
</tr>
</tbody>
</table>

d. Discrete. It may assume 3 values: 0, 1, and 2.

2. a. Let \( x \) = time (in minutes) to assemble the product.

   b. It may assume any positive value: \( x > 0 \).

   c. Continuous

3. Let \( Y \) = position is offered
   \( N \) = position is not offered

   a. \( S = \{(Y,Y,Y), (Y,Y,N), (Y,N,Y), (N,Y,Y), (Y,N,N), (N,Y,N), (N,N,Y), (N,N,N)\} \)

   b. Let \( N \) = number of offers made; \( N \) is a discrete random variable.

   c. 

<table>
<thead>
<tr>
<th>Experimental Outcome</th>
<th>(Y,Y,Y)</th>
<th>(Y,Y,N)</th>
<th>(Y,N,Y)</th>
<th>(N,Y,Y)</th>
<th>(Y,N,N)</th>
<th>(N,Y,N)</th>
<th>(N,N,Y)</th>
<th>(N,N,N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of ( N )</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

4. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

6. a. values: 0,1,2,...,20
   discrete

   b. values: 0,1,2,...
   discrete

   c. values: 0,1,2,...,50
   discrete

d. values: \( 0 \leq x \leq 8 \)
   continuous

e. values: \( x > 0 \)

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
7. a. \( f(x) \geq 0 \) for all values of \( x \).

\[
\sum f(x) = 1 \quad \text{Therefore, it is a proper probability distribution.}
\]

b. Probability \( x = 30 \) is \( f(30) = .25 \)

c. Probability \( x \leq 25 \) is \( f(20) + f(25) = .20 + .15 = .35 \)

d. Probability \( x > 30 \) is \( f(35) = .40 \)

8. a. Let \( x = \) number of operating rooms in use on any given day

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/20 = .15</td>
</tr>
<tr>
<td>2</td>
<td>5/20 = .25</td>
</tr>
<tr>
<td>3</td>
<td>8/20 = .40</td>
</tr>
<tr>
<td>4</td>
<td>4/20 = .20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>

b. 

c. \( f(x) \geq 0 \) for \( x = 1, 2, 3, 4 \).

\[
\sum f(x) = 1
\]

10. a. Senior Executives

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.09</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>0.42</td>
</tr>
<tr>
<td>5</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>

b. Middle Managers

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.04</td>
</tr>
</tbody>
</table>
2  0.10
3  0.12
4  0.46
5  0.28
1.00

c.  \[ P(4 \text{ or } 5) = f(4) + f(5) = 0.42 + 0.41 = 0.83 \]

d.  Probability of very satisfied: 0.28

e.  Senior executives appear to be more satisfied than middle managers. 83% of senior executives have a score of 4 or 5 with 41% reporting a 5. Only 28% of middle managers report being very satisfied.

12.  a.  Yes; \( f(x) \geq 0 \), \( \sum f(x) = 1 \)

b.  \( f(500,000) + f(600,000) = .10 + .05 = .15 \)

c.  \( f(100,000) = .10 \)

14.  a.  \[ f(200) = 1 - f(-100) - f(0) - f(50) - f(100) - f(150) \]

\[ = 1 - .95 = .05 \]

This is the probability MRA will have a $200,000 profit.

b.  \[ P(\text{Profit}) = f(50) + f(100) + f(150) + f(200) \]

\[ = .30 + .25 + .10 + .05 = .70 \]

c.  \[ P(\text{at least 100}) = f(100) + f(150) + f(200) \]

\[ = .25 + .10 + .05 = .40 \]

16.  a.  

<table>
<thead>
<tr>
<th>( y )</th>
<th>( f(y) )</th>
<th>( y f(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.2</td>
<td>.4</td>
</tr>
<tr>
<td>4</td>
<td>.3</td>
<td>1.2</td>
</tr>
<tr>
<td>7</td>
<td>.4</td>
<td>2.8</td>
</tr>
<tr>
<td>8</td>
<td>.1</td>
<td>.8</td>
</tr>
<tr>
<td>10</td>
<td>1.0</td>
<td>5.2</td>
</tr>
</tbody>
</table>

\[ E(y) = \mu = 5.2 \]

b.  

<table>
<thead>
<tr>
<th>( y )</th>
<th>( y - \mu )</th>
<th>( (y - \mu)^2 )</th>
<th>( f(y) )</th>
<th>( (y - \mu)^2 f(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-3.20</td>
<td>10.24</td>
<td>.20</td>
<td>2.048</td>
</tr>
<tr>
<td>4</td>
<td>-1.20</td>
<td>1.44</td>
<td>.30</td>
<td>.432</td>
</tr>
<tr>
<td>7</td>
<td>1.80</td>
<td>3.24</td>
<td>.40</td>
<td>1.296</td>
</tr>
<tr>
<td>8</td>
<td>2.80</td>
<td>7.84</td>
<td>.10</td>
<td>.784</td>
</tr>
</tbody>
</table>

\[ \text{Var}(y) = 4.56 \]

\[ \sigma = \sqrt{4.56} = 2.14 \]
18. a/b/ Owner occupied

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f(x)$</th>
<th>$xf(x)$</th>
<th>$x - \mu$</th>
<th>$(x - \mu)^2$</th>
<th>$(x - \mu)^2f(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.2188</td>
<td>.0000</td>
<td>-1.1825</td>
<td>1.3982</td>
<td>.3060</td>
</tr>
<tr>
<td>1</td>
<td>.5484</td>
<td>.5484</td>
<td>-1.825</td>
<td>.0333</td>
<td>.0183</td>
</tr>
<tr>
<td>2</td>
<td>.1241</td>
<td>.2483</td>
<td>.8175</td>
<td>.6684</td>
<td>.0830</td>
</tr>
<tr>
<td>3</td>
<td>.0489</td>
<td>.1466</td>
<td>1.8175</td>
<td>3.3035</td>
<td>.1614</td>
</tr>
<tr>
<td>4</td>
<td>.0598</td>
<td>.2393</td>
<td>2.8175</td>
<td>7.9386</td>
<td>.4749</td>
</tr>
<tr>
<td>Total</td>
<td>1.0000</td>
<td>1.1825</td>
<td>1.0435</td>
<td>1.0435</td>
<td></td>
</tr>
</tbody>
</table>

Renter occupied

<table>
<thead>
<tr>
<th>$y$</th>
<th>$f(y)$</th>
<th>$yf(y)$</th>
<th>$y - \mu$</th>
<th>$(y - \mu)^2$</th>
<th>$(y - \mu)^2f(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.2497</td>
<td>.0000</td>
<td>-1.2180</td>
<td>1.4835</td>
<td>.3704</td>
</tr>
<tr>
<td>1</td>
<td>.4816</td>
<td>.4816</td>
<td>-2.180</td>
<td>.0475</td>
<td>.0229</td>
</tr>
<tr>
<td>2</td>
<td>.1401</td>
<td>.2801</td>
<td>.7820</td>
<td>.6115</td>
<td>.0856</td>
</tr>
<tr>
<td>3</td>
<td>.0583</td>
<td>.1749</td>
<td>1.7820</td>
<td>3.1755</td>
<td>.1851</td>
</tr>
<tr>
<td>4</td>
<td>.0703</td>
<td>.2814</td>
<td>2.7820</td>
<td>7.7395</td>
<td>.5444</td>
</tr>
<tr>
<td>Total</td>
<td>1.0000</td>
<td>1.2180</td>
<td>1.2085</td>
<td>1.2085</td>
<td></td>
</tr>
</tbody>
</table>

e. The expected number of times that owner-occupied units have a water supply stoppage lasting 6 or more hours in the past 3 months is 1.1825, slightly less than the expected value of 1.2180 for renter-occupied units. And, the variability is somewhat less for owner-occupied units (1.0435) as compared to renter-occupied units (1.2085).

20. a.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f(x)$</th>
<th>$xf(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.85</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>.04</td>
<td>20</td>
</tr>
<tr>
<td>1000</td>
<td>.04</td>
<td>40</td>
</tr>
<tr>
<td>3000</td>
<td>.03</td>
<td>90</td>
</tr>
<tr>
<td>5000</td>
<td>.02</td>
<td>100</td>
</tr>
<tr>
<td>8000</td>
<td>.01</td>
<td>80</td>
</tr>
<tr>
<td>10000</td>
<td>.01</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>430</td>
</tr>
</tbody>
</table>

The expected value of the insurance claim is $430. If the company charges $430 for this type of collision coverage, it would break even.

b. From the point of view of the policyholder, the expected gain is as follows:

Expected Gain = Expected claim payout – Cost of insurance coverage
= $430 - $520 = -$90

The policyholder is concerned that an accident will result in a big repair bill if there is no insurance coverage. So even though the policyholder has an expected annual loss of $90, the insurance is protecting against a large loss.
21. Excel Tables Computations for a, b, and c

<table>
<thead>
<tr>
<th></th>
<th>(x)</th>
<th>(f(x))</th>
<th>(x \cdot f(x))</th>
<th>(x - \mu)</th>
<th>((x - \mu)^2)</th>
<th>((x - \mu)^2 f(x))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.05</td>
<td>0.05</td>
<td>-3.0500</td>
<td>9.3025</td>
<td>0.465125</td>
</tr>
<tr>
<td>Executives</td>
<td>2</td>
<td>0.09</td>
<td>0.18</td>
<td>-2.0500</td>
<td>4.1025</td>
<td>0.376225</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.03</td>
<td>0.09</td>
<td>-1.0500</td>
<td>1.1025</td>
<td>0.033075</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.42</td>
<td>1.68</td>
<td>0.0500</td>
<td>0.0025</td>
<td>0.00105</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.41</td>
<td>2.05</td>
<td>0.0500</td>
<td>0.0025</td>
<td>0.00105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.05</td>
<td></td>
<td>1.2475</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.04</td>
<td>0.04</td>
<td>-2.8400</td>
<td>8.6656</td>
<td>0.322624</td>
</tr>
<tr>
<td>Middle Mgrs</td>
<td>2</td>
<td>0.10</td>
<td>0.2</td>
<td>-1.8400</td>
<td>3.3856</td>
<td>0.33856</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.36</td>
<td>-0.8400</td>
<td>0.7056</td>
<td>0.034672</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.46</td>
<td>1.84</td>
<td>0.1500</td>
<td>0.0256</td>
<td>0.011776</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.28</td>
<td>1.4</td>
<td>1.1500</td>
<td>1.3456</td>
<td>0.376768</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.84</td>
<td></td>
<td>1.1344</td>
</tr>
</tbody>
</table>

a. \(E(x) = \Sigma x f(x) = 0.05(1) + 0.09(2) + 0.03(3) + 0.42(4) + 0.41(5) = 4.05\)

b. \(E(x) = \Sigma x f(x) = 0.04(1) + 0.10(2) + 0.12(3) + 0.46(4) + 0.28(5) = 3.84\)

c. Executives: \(\sigma^2 = \Sigma (x - \mu)^2 f(x) = 1.25\)

Middle Managers: \(\sigma^2 = \Sigma (x - \mu)^2 f(x) = 1.13\)

d. Executives: \(\sigma = 1.12\)

Middle Managers: \(\sigma = 1.07\)

e. The senior executives have a higher average score: 4.05 vs. 3.84 for the middle managers. The executives also have a slightly higher standard deviation.

24. a. Medium \(E(x) = \Sigma x f(x)\)

\[= 50 (.20) + 150 (.50) + 200 (.30) = 145\]

Large: \(E(x) = \Sigma x f(x)\)

\[= 0 (.20) + 100 (.50) + 300 (.30) = 140\]

Medium preferred.

b. Medium

\[
\begin{array}{cccccc}
  x & f(x) & x - \mu & (x - \mu)^2 & (x - \mu)^2 f(x) \\
  50 & .20 & -95 & 9025 & 1805.0 \\
  150 & .50 & 5 & 25 & 12.5 \\
  200 & .30 & 55 & 3025 & 907.5 \\
\end{array}
\]

\(\sigma^2 = 2725.0\)

Large

\[
\begin{array}{cccccc}
  y & f(y) & y - \mu & (y - \mu)^2 & (y - \mu)^2 f(y) \\
  0 & .20 & -140 & 19600 & 3920 \\
  100 & .50 & -40 & 1600 & 800 \\
\end{array}
\]
Medium preferred due to less variance.

25. a. 

\[ E(x) = .2(50) + .5(30) + .3(40) = 37 \]
\[ E(y) = .2(80) + .5(50) + .3(60) = 59 \]
\[ Var(x) = .2(50 - 37)^2 + .5(30 - 37)^2 + .3(40 - 37)^2 = 61 \]
\[ Var(y) = .2(80 - 59)^2 + .5(50 - 59)^2 + .3(60 - 59)^2 = 129 \]

b. 

\[
\begin{array}{c|c}
 x + y & f(x + y) \\
\hline
 130 & .2 \\
 80 & .5 \\
 100 & .3 \\
\end{array}
\]

c. 

\[
\begin{array}{c|c|c|c|c|c}
 x + y & f(x + y) & (x + y)f(x + y) & x + y - E(x + y) & \left[x + y - E(x + y)\right]^2 & f(x + y) \left[x + y - E(x + y)\right]^2 \\
\hline
 130 & .2 & 26 & 34 & 1156 & 231.2 \\
 80 & .5 & 40 & -16 & 256 & 128.0 \\
 100 & .3 & 30 & 4 & 16 & 4.8 \\
\end{array}
\]

\[ E(x + y) = 96 \]
\[ Var(x + y) = 364 \]

d. 

\[ \sigma_{xy} = \frac{Var(x + y) - Var(x) - Var(y)}{2} = \frac{364 - 61 - 129}{2} = 87 \]
\[ Var(x) = 61 \text{ and } Var(y) = 129 \] were computed in part (a), so
\[ \sigma_x = \sqrt{61} = 7.8102 \]
\[ \sigma_y = \sqrt{129} = 11.3578 \]

\[ \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{87}{(7.8102)(11.3578)} = .98 \]

The random variables \( x \) and \( y \) are positively related. Both the covariance and correlation coefficient are positive. Indeed, they are very highly correlated; the correlation coefficient is almost equal to 1.

e. 

\[ Var(x + y) = Var(x) + Var(y) + 2\sigma_{xy} = 61 + 129 + 2(87) = 364 \]
\[ Var(x) + Var(y) = 61 + 129 = 190 \]

The variance of the sum of \( x \) and \( y \) is greater than the sum of the variances by two times the covariance: \( 2(87) = 174 \). The reason it is positive is that, in this case the variables are positively related. Whenever two random variables are positively related, the variance of the sum of the randomly variables will be greater than the sum of the variances of the individual random variables.
26. a. The standard deviation for these two stocks is the square root of the variance.

\[ \sigma_x = \sqrt{\text{Var}(x)} = \sqrt{25} = 5\% \]

\[ \sigma_y = \sqrt{\text{Var}(y)} = 1\% \]

Investments in Stock 1 would be considered riskier than investments in Stock 2 because the standard deviation is higher. Note that if the return for Stock 1 falls $8.45/5 = 1.69$ or more standard deviation below its expected value, an investor in that stock will experience a loss. The return for Stock 2 would have to fall 3.2 standard deviations below its expected value before an investor in that stock would experience a loss.

b. Since \( x \) represents the percent return for investing in Stock 1, the expected return for investing $100 in Stock 1 is $8.45 and the standard deviation is $5.00. So to get the expected return and standard deviation for a $500 investment we just multiply by 5.

Expected return ($500 investment) = 5($8.45) = $42.25  
Standard deviation ($500 investment) = 5($5.00) = $25.00

c. Since \( x \) represents the percent return for investing in Stock 1 and \( y \) represents the percent return for investing in Stock 2, we want to compute the expected value and variance for \( .5x + .5y \).

\[
E(.5x + .5y) = .5E(x) + .5E(y) = .5(8.45) + .5(3.2) = 4.225 + 1.6 = 5.825
\]

\[
\text{Var}(.5x + .5y) = .5^2\text{Var}(x) + .5^2\text{Var}(y) + 2(.5)(.5)\sigma_{xy}
\]

\[
= (.5)^2(25) + (.5)^2(1) + 2(.5)(.5)(-3)
\]

\[= 6.25 + .25 - 1.50 = 5 \]

\[\sigma_{.5x+.5y} = \sqrt{5} = 2.236\]

d. Since \( x \) represents the percent return for investing in Stock 1 and \( y \) represents the percent return for investing in Stock 2, we want to compute the expected value and variance for \( .7x + .3y \).

\[
E(.7x + .3y) = .7E(x) + .3E(y) = .7(8.45) + .3(3.2) = 5.915 + .96 = 6.875
\]

\[
\text{Var}(.7x + .3y) = .7^2\text{Var}(x) + .3^2\text{Var}(y) + 2(.7)(.3)\sigma_{xy}
\]

\[= .7^2(25) + .3^2(1) + 2(.7)(.3)(-3)
\]

\[= 12.25 + .09 - 1.26 = 11.08 \]

\[\sigma_{.7x+.3y} = \sqrt{11.08} = 3.329\]

e. The standard deviations of \( x \) and \( y \) were computed in part (a). The correlation coefficient is given by

\[
\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-3}{(5)(1)} = -.6
\]

There is a fairly strong negative relationship between the variables.
27. a. Dividing each of the frequencies in the table by the total number of restaurants provides the joint probability table below. The bivariate probability for each pair of quality and meal price is shown in the body of the table. This is the bivariate probability distribution. For instance, the probability of a rating of 2 on quality and a rating of 3 on meal price is given by \( f(2, 3) = .18 \). The marginal probability distribution for quality, \( x \), is in the rightmost column. The marginal probability for meal price, \( y \), is in the bottom row.

<table>
<thead>
<tr>
<th>Quality (( x ))</th>
<th>Meal Price (( y ))</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.14</td>
<td>0.13</td>
<td>0.01</td>
<td>0.28</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.11</td>
<td>0.21</td>
<td>0.18</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.26</td>
<td>0.39</td>
<td>0.35</td>
<td>1.00</td>
</tr>
</tbody>
</table>

b. \[ E(x) = 1(.28) + 2(.50) + 3(.22) = 1.94 \]

\[ \text{Var}(x) = .28(1 - 1.94)^2 + .50(2 - 1.94)^2 + .22(3 - 1.94)^2 = .4964 \]

c. \[ E(y) = 1(.26) + 2(.39) + 3(.35) = 2.09 \]

\[ \text{Var}(y) = .26(1 - 2.09)^2 + .39(2 - 2.09)^2 + .35(3 - 2.09)^2 = .6019 \]

d. \[ \sigma_{xy} = \frac{\text{Var}(x + y) - \text{Var}(x) - \text{Var}(y)}{2} = \frac{[1.6691 - .4964 - .6019]}{2} = .2854 \]

Since, the covariance \( \sigma_{xy} = .2854 \) is positive we can conclude that as the quality rating goes up, the meal price goes up. This is as we would expect.

e. \[ \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{.2854}{\sqrt{.4964} \sqrt{.6019}} = .5221 \]

With a correlation coefficient of .5221 we would call this a moderately positive relationship. It is not likely to find a low cost restaurant that is also high quality. But, it is possible. There are 3 of them leading to \( f(3,1) = .01 \).

28. a. Marginal distribution of Direct Labor Cost

<table>
<thead>
<tr>
<th>( y )</th>
<th>( f(y) )</th>
<th>( yf(y) )</th>
<th>( y - E(y) )</th>
<th>( (y - E(y))^2 )</th>
<th>( (y - E(y))^2f(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>.3</td>
<td>12.9</td>
<td>-2.3</td>
<td>5.29</td>
<td>1.587</td>
</tr>
<tr>
<td>45</td>
<td>.4</td>
<td>18</td>
<td>-.3</td>
<td>.09</td>
<td>.036</td>
</tr>
<tr>
<td>48</td>
<td>.3</td>
<td>14.4</td>
<td>2.7</td>
<td>7.29</td>
<td>2.187</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>45.3</td>
<td></td>
<td>3.81</td>
<td></td>
</tr>
</tbody>
</table>

\[ E(y) = 45.3 \]

\[ \sigma_y = 1.95 \]

b. Marginal distribution of Parts Cost

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
<th>( xf(x) )</th>
<th>( x - E(x) )</th>
<th>( (x - E(x))^2 )</th>
<th>( (x - E(x))^2f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>.45</td>
<td>38.25</td>
<td>-5.5</td>
<td>30.25</td>
<td>13.6125</td>
</tr>
<tr>
<td>95</td>
<td>.55</td>
<td>52.25</td>
<td>4.5</td>
<td>20.25</td>
<td>11.1375</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>90.5</td>
<td></td>
<td>24.75</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Var}(x) = .2854 \]

\[ \text{Var}(y) = 1.95 \]
c. Let $z = x + y$ represent total manufacturing cost (direct labor + parts).

<table>
<thead>
<tr>
<th>$z$</th>
<th>$f(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>.05</td>
</tr>
<tr>
<td>130</td>
<td>.20</td>
</tr>
<tr>
<td>133</td>
<td>.20</td>
</tr>
<tr>
<td>138</td>
<td>.25</td>
</tr>
<tr>
<td>140</td>
<td>.20</td>
</tr>
<tr>
<td>143</td>
<td>.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$zf(z)$</th>
<th>$z - E(z)$</th>
<th>$(z - E(z))^2$</th>
<th>$(z - E(z))^2f(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>-7.8</td>
<td>60.84</td>
<td>3.042</td>
</tr>
<tr>
<td>26</td>
<td>-5.8</td>
<td>33.64</td>
<td>6.728</td>
</tr>
<tr>
<td>26.6</td>
<td>-2.8</td>
<td>7.84</td>
<td>1.568</td>
</tr>
<tr>
<td>34.5</td>
<td>2.2</td>
<td>4.84</td>
<td>1.21</td>
</tr>
<tr>
<td>28</td>
<td>4.2</td>
<td>17.64</td>
<td>3.528</td>
</tr>
<tr>
<td>14.3</td>
<td>7.2</td>
<td>51.84</td>
<td>5.184</td>
</tr>
</tbody>
</table>

$E(z) = 135.8$  
$Var(z) = 21.26$  
$\sigma_z = 4.61$

e. To determine if $x = \text{parts cost}$ and $y = \text{direct labor cost}$ are independent, we need to compute the covariance $\sigma_{xy}$.

$$\sigma_{xy} = \frac{(\text{Var}(x + y) - \text{Var}(x) - \text{Var}(y))/2}{\text{Var}(x)\text{Var}(y)} = \frac{(21.26 - 24.75 - 3.81)/2}{-3.65} = 4.61$$

Since the covariance is not equal to zero, we can conclude that direct labor cost is not independent of parts cost. Indeed, they are negatively correlated. When parts cost goes up, direct labor cost goes down. Maybe the parts costing $95 come from a different manufacturer and are higher quality. Working with higher quality parts may reduce labor costs.

f. The expected manufacturing cost for 1500 printers is

$$E(1500z) = 1500E(z) = 1500(135.8) = 203,700$$

The total manufacturing costs of $198,350 are less than we would have expected. Perhaps as more printers were manufactured there was a learning curve and direct labor costs went down.

30. a. Let $x = \text{percentage return for S&P 500}$  
$y = \text{percentage return for Core Bond fund}$  
$z = \text{percentage return for REITs}$

The formula for computing the correlation coefficient is given by
\[ \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \]

In this case, we know the correlation coefficients and the 3 standard deviations, so we want to rearrange the correlation coefficient formula to find the covariances.

S&P 500 and REITs: \( \sigma_{x_2} = \rho_{x_2} \sigma_x \sigma_y = (.74)(19.45)(23.17) = 333.486 \)

Core Bonds and REITs: \( \sigma_{y_2} = 2.13(23.17) = (-.04)(2.13)(23.17) = -1.974 \)

b. Letting \( r = \) portfolio percentage return, we have \( r = .5x + .5y \). The expected return for a portfolio with 50% invested in the S&P 500 and 50% invested in REITs is

\[ E(r) = .5E(x) + .5E(z) = (.5)5.04% + (.5)13.07% = 9.055% \]

We are given \( \sigma_x \) and \( \sigma_z \), so \( Var(x) = 19.45^2 = 378.3025 \) and \( Var(z) = 23.17^2 = 536.8489 \). We can now compute

\[ Var(.5x + .5z) = .5^2Var(x) + .5^2Var(z) + 2(.5)(.5)(\sigma_{xz}) \]

\[ = .25(378.3025) + .25(536.8489) + .5(333.486) \]

\[ = 395.53 \]

Then \( \sigma_{xy} = \sqrt{395.53} = 19.89 \)

So, the expected return for our portfolio is 9.055% and the standard deviation is 19.89%.

c. Letting \( r = \) portfolio percentage return, we have \( r = .5y + .5z \). The expected return for a portfolio with 50% invested in Core Bonds and 50% invested in REITs is

\[ E(r) = .5E(y) + .5E(z) = (.5)5.78% + (.5)13.07% = 9.425% \]

We are given \( \sigma_y \) and \( \sigma_z \), so \( Var(y) = 2.13^2 = 4.5369 \) and \( Var(z) = 23.17^2 = 536.8489 \). We can now compute

\[ Var(.5y + .5z) = .5^2Var(y) + .5^2Var(z) + 2(.5)(.5)(\sigma_{yz}) \]

\[ = .25(4.5369) + .25(536.8489) + .5(-.04) \]

\[ = 135.33 \]

Then \( \sigma_{yz} = \sqrt{135.33} = 11.63 \)

So, the expected return for our portfolio is 9.425% and the standard deviation is 11.63%.

d. Letting \( r = \) portfolio percentage return, we have \( r = .8y + .2z \). The expected return for a portfolio with 80% invested in Core Bonds and 20% invested in REITs is

\[ E(r) = .8E(y) + .2E(z) = (.8)5.78% + (.2)13.07% = 7.238% \]
From part (c) above, we have \( \text{Var}(y) = 4.5369 \) and \( \text{Var}(z) = 536.8489 \). We can now compute

\[
\text{Var}(0.8y + 0.2z) = 0.8^2 \text{Var}(y) + 0.2^2 \text{Var}(z) + 2 \cdot 0.8 \cdot 0.2 \cdot (\sigma_{yz})
\]

\[
= 0.64(4.5369) + 0.04(536.8489) + 0.32(-0.04)
\]

\[
= 24.36
\]

Then \( \sigma_{yz} = \sqrt{24.36} = 4.94 \)

So, the expected return for our portfolio is 7.238\% and the standard deviation is 4.94\%.

e. The expected returns and standard deviations for the 3 portfolios are summarized below.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Expected Return (%)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% S&amp;P 500 &amp; 50% REITs</td>
<td>9.055</td>
<td>19.89</td>
</tr>
<tr>
<td>50% Core Bonds &amp; 50% REITs</td>
<td>9.425</td>
<td>11.63</td>
</tr>
<tr>
<td>80% Core Bonds &amp; 20% REITs</td>
<td>7.238</td>
<td>4.94</td>
</tr>
</tbody>
</table>

The portfolio from part (c) involving 50% Core Bonds and 50% REITS has the highest return. Using the standard deviation as a measure of risk, it also has less risk than the portfolio from part (b) involving 50% invested in an S&P 500 index fund and 50% invested in REITs. So the portfolio from part (b) would not be recommended for either type of investor.

The portfolio from part (d) involving 80% in Core Bonds and 20% in REITs has the lowest standard deviation and thus lesser risk than the portfolio in part (c). We would recommend the portfolio consisting of 50% Core Bonds and 50% REITs for the aggressive investor because of its higher return and moderate amount of risk.

We would recommend the portfolio consisting of 80% Core Bonds and 20% REITS to the conservative investor because of its low risk and moderate return.

31. a.

b. \( f(1) = \binom{2}{1} (0.4)^1 (0.6)^1 = \frac{2!}{1!1!} (0.4)(0.6) = 0.48 \)

Using Excel: BINOM.DIST(1,2,.4,FALSE) = 0.48

c. \( f(0) = \binom{2}{0} (0.4)^0 (0.6)^2 = \frac{2!}{0!2!} (1)(0.36) = 0.36 \)
Using Excel: BINOM.DIST(0,2,.4,FALSE) = .36

d. \( f(2) = \binom{2}{2} (0.4)^2 (0.6)^0 = \frac{2!}{2!0!} (0.16)(1) = .16 \)

Using Excel: BINOM.DIST(2,2,.4,FALSE) = .16

e. \( P(x \geq 1) = f(1) + f(2) = .48 + .16 = .64 \)

f. \( E(x) = n p = 2 (0.4) = .8 \)

\( \text{Var}(x) = np(1-p) = 2 (0.4)(0.6) = .48 \)

\( \sigma = \sqrt{0.48} = .6928 \)

32. a. \( f(0) = \text{BINOM.DIST}(0,10,.1,FALSE) = .3487 \)

b. \( f(2) = \text{BINOM.DIST}(2,10,.1,FALSE) = .1937 \)

c. \( P(x \leq 2) = f(0) + f(1) + f(2) = .3487 + .3874 + .1937 = \text{BINOM.DIST}(2,10,.1,TRUE) = .9298 \)

d. \( P(x \geq 1) = 1 - f(0) = 1 - \text{BINOM.DIST}(0,10,.1,FALSE) = 1 - .3487 = .6513 \)

e. \( E(x) = np = 10 (.1) = 1 \)

f. \( \text{Var}(x) = np(1-p) = 10 (.1)(.9) = .9 \)

\( \sigma = \sqrt{0.9} = .9487 \)

34. a. Yes. Since the teenagers are selected randomly, \( p \) is the same from trial to trial and the trials are independent. The two outcomes per trial are use Pandora Media Inc.’s online radio service or do not use Pandora Media Inc.’s online radio service.

\( \text{Binomial} \ \ n = 10 \text{ and } p = .35 \)

\[ f(x) = \frac{10!}{x!(10-x)!} (.35)^x (1-.35)^{10-x} \]

b. \( f(0) = \frac{10!}{0!(10-0)!} (.35)^0 (.65)^{10-0} = .0135 \ \text{ OR } \ \text{BINOM.DIST}(0,10,.35,FALSE) = .0135 \)

c. \( f(4) = \frac{10!}{4!(10-4)!} (.35)^4 (.65)^{10-4} = .2377 \ \text{ OR } \ \text{BINOM.DIST}(4,10,.35,FALSE) = .2377 \)

d. Probability \( (x \geq 2) = 1 - f(0) - f(1) \)

From part (b), \( f(0) = .0135 \)

\[ f(1) = \frac{10!}{1!(10-1)!} (.35)^1 (.65)^{10-1} = .0725 \]

Probability \( (x \geq 2) = 1 - f(0) - f(1) = 1 - (.0135 + .0725) = .9140 \)
36. a. Probability of a defective part being produced must be .03 for each part selected; parts must be selected independently.

b. Let:  
\( D = \) defective  
\( G = \) not defective

<table>
<thead>
<tr>
<th>1st part</th>
<th>2nd part</th>
<th>Number Defective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(D, D) 2</td>
</tr>
<tr>
<td>D</td>
<td>G</td>
<td>(D, G) 1</td>
</tr>
<tr>
<td>G</td>
<td>D</td>
<td>(G, D) 1</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>(G, G) 0</td>
</tr>
</tbody>
</table>

\(f\) outcomes result in exactly one defect.

d. \(P(\text{no defects}) = (.97)(.97) = .9409\)  
\(P(\text{1 defect}) = 2(.03)(.97) = .0582\)  
\(P(\text{2 defects}) = (.03)(.03) = .0009\)

38. a. .90

b. \(P(\text{at least 1}) = f(1) + f(2)\)

\[f(1) = \frac{2!}{1!1!}(.9)^1(.1)^1\]

\[= 2(.9)(.1) = .18\]

\(f(2) = \frac{2!}{2!0!}(.9)^0(.1)^2\)

\[= 1(.81)(1) = .81\]

\[\therefore P(\text{at least 1}) = .18 + .81 = .99\]

Alternatively

\(P(\text{at least 1}) = 1 - f(0)\)
Chapter 5

\[ f(0) = \frac{2!}{0! \cdot 2!} (.9)^0 (.1)^2 = .01 \]

Therefore, \( P(\text{at least 1}) = 1 - .01 = .99 \) OR \( 1 - \text{BINOM.DIST}(0, 2, .9, \text{FALSE}) = .99 \)

c. \( P(\text{at least 1}) = 1 - f(0) \)

\[ f(0) = \frac{3!}{0! \cdot 3!} (.9)^0 (.1)^3 = .001 \]

Therefore, \( P(\text{at least 1}) = 1 - .001 = .999 \) OR \( 1 - \text{BINOM.DIST}(0, 3, .9, \text{FALSE}) = .999 \)

d. Yes; \( P(\text{at least 1}) \) becomes very close to 1 with multiple systems and the inability to detect an attack would be catastrophic.

40. a. Yes. Since the 18- to 34-year olds living with their parents are selected randomly, \( p \) is the same from trial to trial and the trials are independent. The two outcomes per trial are contribute to household expenses or do not contribute to household expenses.

Binomial \( n = 15 \) and \( p = .75 \)

\[ f(x) = \frac{15!}{x!(15-x)!} (.75)^x (1-.75)^{15-x} \]

b. The probability that none of the fifteen contribute to household expenses is

\[ f(0) = \frac{15!}{0!(15-0)!} (.75)^0 (1-.75)^{15-0} = .0000 \]

OR \( \text{BINOM.DIST}(0, 15, .75, \text{FALSE}) = .0000 \)

Obtaining a sample result that shows that none of the fifteen contributed to household expenses is so unlikely you would have to question whether the 75% value reported by the Pew Research Center is accurate.

c. Probability of at least ten = \( f(10) + f(11) + f(12) + f(13) + f(14) + f(15) \)

Using binomial tables

\[
\text{Probability} = .1651 + .2252 + .2252 + .1559 + .0668 + .0134 = .8516 \\
\text{OR} \quad 1 - \text{BINOM.DIST}(9, 15, .75, \text{TRUE}) = .8516
\]

42. a. \( f(4) = \frac{20!}{4!(20-4)!} (.30)^4 (.70)^{20-4} = .1304 \)

OR \( \text{BINOM.DIST}(4, 20, .3, \text{FALSE}) = .1304 \)

b. Probability \( (x \geq 2) = 1 - f(0) - f(1) \)

\[ f(0) = \frac{20!}{0!(20-0)!} (.30)^0 (.70)^{20-0} = .0008 \]

\[ f(1) = \frac{20!}{1!(20-1)!} (.30)^1 (.70)^{20-1} = .0068 \]
Probability \((x \geq 2) = 1 - f(0) - f(1) = 1 - (.0008 + .0068) = .9924\)

OR \(1 - \text{BINOM.DIST}(1,20,.3,\text{TRUE}) = .9924\)

c. \(E(x) = n p = 20(.30) = 6\)

d. \(Var(x) = n p (1 - p) = 20(.30)(1-.30) = 4.2\)

\(\sigma = \sqrt{4.2} = 2.0494\)

44. a. \(f(x) = \frac{3^x e^{-3}}{x!}\)

b. \(f(2) = \frac{3^2 e^{-3}}{2!} = \frac{9(.0498)}{2} = .2240\)

Using Excel: \(\text{POISSON.DIST}(2,3,\text{FALSE}) = .2240\)

c. \(f(1) = \frac{3^1 e^{-3}}{1!} = 3(.0498) = .1494\)

Using Excel: \(\text{POISSON.DIST}(1,3,\text{FALSE}) = .1494\)

d. \(P(x \geq 2) = 1 - f(0) - f(1) = 1 - .0498 - .1494 = .8009\)

Using Excel: \(1 - \text{POISSON.DIST}(1,3,\text{TRUE}) = .8009\)

45. a. \(f(x) = \frac{2^x e^{-2}}{x!}\)

b. \(\mu = 6\) for 3 time periods

c. \(f(x) = \frac{6^x e^{-6}}{x!}\)

d. \(f(2) = \frac{2^2 e^{-2}}{2!} = \frac{4(.1353)}{2} = .2707\)

Using Excel: \(\text{POISSON.DIST}(2,2,\text{FALSE}) = .2707\)

e. \(f(6) = \frac{6^6 e^{-6}}{6!} = .1606\)

Using Excel: \(\text{POISSON.DIST}(6,6,\text{FALSE}) = .1606\)

f. \(f(5) = \frac{4^5 e^{-4}}{5!} = .1563\)

Using Excel: \(\text{POISSON.DIST}(5,4,\text{FALSE}) = .1563\)
46. a. \( \mu = 48 \frac{5}{60} = 4 \) per 5 minutes

\[
f(3) = \frac{4^3 e^{-4}}{3!} = .1954 = \text{POISSON.DIST}(3,4,\text{FALSE}) = .1954
\]

b. \( \mu = 48 \frac{15}{60} = 12 \) per 15 minutes

\[
f(10) = \frac{12^{10} e^{-12}}{10!} = \text{POISSON.DIST}(10,12,\text{FALSE}) = .1048
\]

c. \( \mu = 48 \frac{5}{60} = 4 \) I expect 4 callers to be waiting after 5 minutes.

\[
f(0) = \frac{4^0 e^{-4}}{0!} = \text{POISSON.DIST}(0,4,\text{FALSE}) = .0183
\]

The probability none will be waiting after 5 minutes is .0183.

d. \( \mu = 48 \frac{3}{60} = 2.4 \) per 3 minutes

\[
f(0) = \frac{2.4^0 e^{-2.4}}{0!} = \text{POISSON.DIST}(0,2.4,\text{FALSE}) = .0907
\]

The probability of no interruptions in 3 minutes is .0907.

48. a. For a 15-minute period the mean is \( 14.4/4 = 3.6 \)

\[
f(0) = \frac{3.6^0 e^{-3.6}}{0!} = e^{-3.6} = .0273 \quad \text{OR} \quad \text{POISSON.DIST}(0,3.6,\text{FALSE}) = .0273
\]

b. probability = \( 1 - f(0) = 1 - .2073 = .9727 \)

c. probability = \( 1 - [f(0) + f(1) + f(2) + f(3)] \)

\[
= 1 - [.0273 + .0984 + .1771 + .2125] = .4848
\]

\[
\text{OR} \quad 1 - \text{POISSON.DIST}(3,3.6,\text{TRUE}) = .4848
\]

Note: The value of \( f(0) \) was computed in part (a); a similar procedure was used to compute the probabilities for \( f(1), f(2), \) and \( f(3) \).

50. a. \( \mu = 18/30 = .6 \) per day during June

\[
f(0) = \frac{.6^0 e^{-0.6}}{0!} = .5488 \quad \text{OR} \quad \text{POISSON.DIST}(0,.6,\text{FALSE}) = .5488
\]

b. \( f(1) = \frac{.6^1 e^{-0.6}}{1!} = .3293 \quad \text{OR} \quad \text{POISSON.DIST}(1,.6,\text{FALSE}) = .3293
\]

c. \( P(\text{More than 1}) = 1 - f(0) - f(1) = 1 - .5488 - .3293 = .1219 \)

\[
\text{OR} \quad 1 - \text{POISSON.DIST}(1,.6,\text{TRUE}) = .1219
\]
52. All parts involve the hypergeometric distribution with \(N=10, r=3\)

\[
f(n) = \frac{\binom{r}{n} \binom{N-r}{x-n}}{\binom{N}{n}} = \frac{\frac{r!}{n!(r-n)!} \cdot \frac{(N-r)!}{(x-n)!(N-r-x)!}}{\frac{N!}{n!(N-n)!}}
\]

\(f(1) = \frac{\binom{3}{4} \binom{10-3}{4-1}}{\binom{10}{4}} = \frac{\frac{3!}{1!2!} \cdot \frac{7!}{3!4!} \cdot \frac{10!}{4!6!}}{210} = .50 \quad n=4, x=1
\]

Using Excel: HYPGEOM.DIST(1,4,3,10,FALSE) = .5000

\(f(2) = \frac{\binom{3}{2} \binom{10-3}{2-2}}{\binom{10}{2}} = \frac{\frac{3!}{2!1!} \cdot \frac{7!}{2!5!} \cdot \frac{10!}{4!6!}}{45} = .0667 \quad n=2, x=2
\]

Using Excel: HYPGEOM.DIST(2,2,3,10,FALSE) = .0667

\(f(0) = \frac{\binom{3}{0} \binom{10-3}{0-0}}{\binom{10}{2}} = \frac{\frac{3!}{0!3!} \cdot \frac{7!}{0!7!} \cdot \frac{10!}{4!6!}}{45} = .4667 \quad n=2, x=0
\]

Using Excel: HYPGEOM.DIST(0,2,3,10,FALSE) = .4667

\(f(2) = \frac{\binom{3}{2} \binom{10-3}{4-2}}{\binom{10}{4}} = \frac{\frac{3!}{2!1!} \cdot \frac{7!}{2!5!} \cdot \frac{10!}{4!6!}}{210} = .30 \quad n=4, x=2
\]

Using Excel: HYPGEOM.DIST(2,4,3,10,FALSE) = .3000

e. The scenario of \(n=4, x=4\) is not possible with \(r=3\) because it is not possible to have 4 actual successes \((x)\) out of 3 possible successes \((r)\).

54. Hypergeometric Distribution with \(N = 10\) and \(r = 7\)

\(f(2) = \frac{\binom{7}{2} \binom{3}{1}}{\binom{10}{3}} = \frac{21 \cdot 3}{120} = .5250 \quad n=3, x=2 \quad \text{OR} \quad \text{HYPGEOM.DIST}(2,3,7,10,FALSE) = .5250
\]

b. Compute the probability that 3 prefer shopping online.

\(f(3) = \frac{\binom{7}{3} \binom{3}{0}}{\binom{10}{3}} = \frac{35 \cdot 1}{120} = .2917 \quad n=3, x=3 \quad \text{OR} \quad \text{HYPGEOM.DIST}(3,3,7,10,FALSE) = .2917
\]

\[P(\text{majority prefer shopping online}) = f(2) + f(3) = .5250 + .2917 = .8167\]
56. \( N = 60 \quad n = 10 \)

a. \( r = 20 \quad x = 0 \)

\[
f(0) = \frac{\binom{20}{0} \binom{40}{10}}{\binom{60}{10}} = \frac{(1) \left( \frac{40!}{10!30!} \right)}{\frac{60!}{10!50!}} = \text{HYPGEOM.DIST}(0,10,20,60,\text{FALSE}) = 0.0112
\]

b. \( r = 20 \quad x = 1 \)

\[
f(1) = \frac{\binom{20}{1} \binom{40}{9}}{\binom{60}{10}} = \text{HYPGEOM.DIST}(1,10,20,60,\text{FALSE}) = 0.0725
\]

c. \( 1 - f(0) - f(1) = 1 - 0.0112 - 0.0725 = 0.9163 \approx 0.92 \)

d. Same as the probability one will be from Hawaii. In part b that was found to equal approximately .07. This is also shown with the hypergeometric distribution with \( N=60, \ r=40, \ n=10, \) and \( x=9 \)

\[
\text{OR} \quad \text{HYPGEOM.DIST}(9,10,40,60,\text{FALSE}) = 0.0725
\]

58. Let \( x \) be the number of boxes in the sample that show signs of spoilage. This is a hypergeometric random variable with \( N = 100 \) and \( r = 8 \). For \( n = 10 \) and \( x = 2 \) we have:

\[
f(2) = \frac{\binom{8}{2} \binom{92}{8}}{\binom{100}{10}} = \frac{\frac{8!}{2!6!} \left( \frac{92!}{8!84!} \right)}{\frac{100!}{10!90!}} = \text{HYPGEOM.DIST}(2,10,8,100,\text{FALSE}) = 0.1506
\]

60. a.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.150</td>
</tr>
<tr>
<td>2</td>
<td>0.050</td>
</tr>
<tr>
<td>3</td>
<td>0.075</td>
</tr>
<tr>
<td>4</td>
<td>0.050</td>
</tr>
<tr>
<td>5</td>
<td>0.125</td>
</tr>
<tr>
<td>6</td>
<td>0.050</td>
</tr>
<tr>
<td>7</td>
<td>0.100</td>
</tr>
<tr>
<td>8</td>
<td>0.125</td>
</tr>
<tr>
<td>9</td>
<td>0.125</td>
</tr>
<tr>
<td>10</td>
<td>0.150</td>
</tr>
<tr>
<td>Total</td>
<td>1.000</td>
</tr>
</tbody>
</table>

b. Probability of outstanding service is \( 0.125 + 0.150 = 0.275 \)

c.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
<th>( xf(x) )</th>
<th>( x - \mu )</th>
<th>( (x - \mu)^2 )</th>
<th>( (x - \mu)^2 f(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.150</td>
<td>0.150</td>
<td>-4.925</td>
<td>24.2556</td>
<td>3.6383</td>
</tr>
</tbody>
</table>
### Discrete Probability Distributions

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.050</td>
<td>.100</td>
<td>-3.925</td>
</tr>
<tr>
<td>3</td>
<td>.075</td>
<td>.225</td>
<td>-2.925</td>
</tr>
<tr>
<td>4</td>
<td>.050</td>
<td>.200</td>
<td>-1.925</td>
</tr>
<tr>
<td>5</td>
<td>.125</td>
<td>.625</td>
<td>-.925</td>
</tr>
<tr>
<td>6</td>
<td>.050</td>
<td>.300</td>
<td>.075</td>
</tr>
<tr>
<td>7</td>
<td>.100</td>
<td>.700</td>
<td>1.075</td>
</tr>
<tr>
<td>8</td>
<td>.125</td>
<td>1.000</td>
<td>2.075</td>
</tr>
<tr>
<td>9</td>
<td>.125</td>
<td>1.125</td>
<td>3.075</td>
</tr>
<tr>
<td>10</td>
<td>.150</td>
<td>1.500</td>
<td>4.075</td>
</tr>
<tr>
<td>Total</td>
<td>1.000</td>
<td>5.925</td>
<td>9.6694</td>
</tr>
</tbody>
</table>

\[ E(x) = 5.925 \text{ and } Var(x) = 9.6694 \]

The probability of a new car dealership receiving an outstanding wait-time rating is \( \frac{2}{7} = .2857 \). For the remaining \( 40 - 7 = 33 \) service providers, 9 received an outstanding rating; this corresponds to a probability of \( \frac{9}{33} = .2727 \). For these results, there does not appear to be much difference between the probability that a new car dealership is rated outstanding compared to the same probability for other types of service providers.

62. a. There are 600 observations involving the two variables. Dividing the entries in the table shown by 600 and summing the rows and columns we obtain the following.

<table>
<thead>
<tr>
<th>Snacks (x)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.0</td>
<td>.1</td>
<td>.03</td>
<td>.13</td>
</tr>
<tr>
<td>1</td>
<td>.4</td>
<td>.15</td>
<td>.05</td>
<td>.6</td>
</tr>
<tr>
<td>2</td>
<td>.2</td>
<td>.05</td>
<td>.02</td>
<td>.27</td>
</tr>
<tr>
<td>Total</td>
<td>.6</td>
<td>.3</td>
<td>.1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The entries in the body of the table are the bivariate or joint probabilities for \( x \) and \( y \). The entries in the rightmost (Total) column are the marginal probabilities for \( x \) and the entries in the bottom (Total) row are the marginal probabilities for \( y \).

The probability of a customer purchasing 1 item of reading materials and 2 snack items is given by \( f(x = 1, y = 2) = .05 \).

The probability of a customer purchasing 1 snack item only is given by \( f(x = 1, y = 0) = .40 \).

The probability \( f(x = 0, y = 0) = 0 \) because the point of sale terminal is only used when someone makes a purchase.

b. The marginal probability distribution of \( x \) along with the calculation of the expected value and variance is shown below.

\[
\begin{array}{c|c|c|c|c|c}
  x & f(x) & xf(x) & x - E(x) & (x - E(x))^2 & (x-E(x))^2f(x) \\
  \hline
  0 & 0.13 & 0 & -1.14 & 1.2996 & 0.1689 \\
  1 & 0.60 & 0.6 & -0.14 & 0.0196 & 0.0118 \\
  2 & 0.27 & 0.54 & 0.86 & 0.7396 & 0.1997 \\
  \hline
  \text{Total} & 1.14 & & & & 0.3804 \\
\end{array}
\]

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
We see that $E(x) = 1.14$ snack items and $Var(x) = .3804$.

c. The marginal probability distribution of $y$ along with the calculation of the expected value and variance is shown below.

<table>
<thead>
<tr>
<th>$y$</th>
<th>$f(y)$</th>
<th>$yf(y)$</th>
<th>$y - E(y)$</th>
<th>$(y - E(y))^2$</th>
<th>$(y - E(y))^2f(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.60</td>
<td>0</td>
<td>-0.5</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.3</td>
<td>0.5</td>
<td>0.25</td>
<td>0.075</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.2</td>
<td>1.5</td>
<td>2.25</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>

We see that $E(y) = .50$ reading materials and $Var(y) = .45$.

d. The probability distribution of $t = x + y$ is shown below along with the calculation of its expected value and variance.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$f(t)$</th>
<th>$tf(t)$</th>
<th>$t - E(t)$</th>
<th>$(t - E(t))^2$</th>
<th>$(t - E(t))^2f(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.5</td>
<td>-0.64</td>
<td>0.4096</td>
<td>0.2048</td>
</tr>
<tr>
<td>2</td>
<td>0.38</td>
<td>0.76</td>
<td>0.36</td>
<td>0.1296</td>
<td>0.0492</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
<td>0.3</td>
<td>1.36</td>
<td>1.8496</td>
<td>0.1850</td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>0.08</td>
<td>2.36</td>
<td>5.5696</td>
<td>0.1114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.64</td>
<td>0.5504</td>
<td></td>
</tr>
</tbody>
</table>

We see that the expected number of items purchased is $E(t) = 1.64$ and the variance in the number of purchases is $Var(t) = .5504$.

e. From part (b), $Var(x) = .3804$. From part (c), $Var(y) = .45$. And from part (d), $Var(x + y) = Var(t) = .5504$. Therefore,

$$\sigma_{xy} = \frac{[Var(x + y) - Var(x) - Var(y)]}{2}$$

$$= \frac{(0.5504 - 0.3804 - 0.4500)}{2}$$

$$= -0.14$$

To compute the correlation coefficient, we must first obtain the standard deviation of $x$ and $y$.

$$\sigma_x = \sqrt{Var(x)} = \sqrt{.3804} = .6168$$

$$\sigma_y = \sqrt{Var(y)} = \sqrt{.45} = .6708$$

So the correlation coefficient is given by
The relationship between the number of reading materials purchased and the number of snacks purchased is negative. This means that the more reading materials purchased the fewer snack items purchased and vice versa.

64. a. \( n = 20, p = .53 \) and \( x = 3 \)

\[
f(3) = \binom{20}{3} (.53)^3 (.47)^2 = \text{BINOM.DIST}(3,20,.53,\text{FALSE}) = .0005
\]

b. \( n = 20, p = .28 \) and \( x = 0, 1, 2, 3, 4, 5 \)

\[
P(x \leq 5) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) = \text{BINOM.DIST}(5,20,.28,\text{TRUE}) = .4952
\]

c. \( E(x) = np = 2000(.49) = 980 \)

The expected number who would find it very hard to give up their smartphone is 980.

d. \( E(x) = np = 2000(.36) = 720 \)

The expected number who would find it very hard to give up their E-mail is 720.

\[
\sigma^2 = np(1 - p) = 2000(.36)(.64) = 460.8
\]

\[
\sigma = \sqrt{460.8} = 21.4663
\]

66. Since the shipment is large we can assume that the probabilities do not change from trial to trial and use the binomial probability distribution.

a. \( n = 5 \)

\[
f(0) = \binom{5}{0} (0.01)^0 (0.99)^5 = \text{BINOM.DIST}(0,5,.01,\text{FALSE}) = .9510
\]

b. \( f(1) = \binom{5}{1} (0.01)^1 (0.99)^4 = \text{BINOM.DIST}(1,5,.01,\text{FALSE}) = .0480
\]

c. \( 1 - f(0) = 1 - .9510 = .0490 \)

d. No, the probability of finding one or more items in the sample defective when only 1% of the items in the population are defective is small (only .0490). I would consider it likely that more than 1% of the items are defective.

68. a. \( E(x) = 200(.235) = 47 \)

b. \( \sigma = \sqrt{np(1 - p)} = \sqrt{200(.235)(.765)} = 5.9962 \)
c. For this situation $p = .765$ and $(1-p) = .235$; but the answer is the same as in part (b). For a binomial probability distribution, the variance for the number of successes is the same as the variance for the number of failures. Of course, this also holds true for the standard deviation.

70. $\mu = 1.5$

Probability of 3 or more breakdowns is $1 - [f(0) + f(1) + f(2)]$.

\[
1 - [f(0) + f(1) + f(2)]
\]
\[
= 1 - [.2231 + .3347 + .2510]
\]
\[
= 1 - .8088 = .1912
\]

Using Excel: $1 - \text{POISSON.DIST}(2,1.5,\text{TRUE}) = .1912$

72. a. $f(3) = \frac{3^3e^{-3}}{3!} = \text{POISSON.DIST}(3,3,\text{FALSE}) = .2240$

b. $f(3) + f(4) + \cdots = 1 - [f(0) + f(1) + f(2)]$

\[
f(0) = \frac{3^0e^{-3}}{0!} = e^{-3} = .0498
\]

Similarly, $f(1) = .1494, f(2) = .2240$

\[
\therefore 1 - [.0498 + .1494 + .2240] = .5768
\]

Using Excel: $1 - \text{POISSON.DIST}(2,3,\text{TRUE}) = .5768$
Chapter 6
Continuous Probability Distributions

Solutions:

1. a. 

![Graph of a function](image1)

\[ f(x) \]

b. \[ P(x = 1.25) = 0 \]. The probability of any single point is zero since the area under the curve above any single point is zero.

c. \[ P(1.0 \leq x \leq 1.25) = \frac{1}{10} \times .25 = .25 \]

d. \[ P(1.20 < x < 1.5) = \frac{1}{10} \times .30 = .30 \]

2. a. 

![Graph of a function](image2)

\[ f(x) \]

b. \[ P(x < 15) = \frac{1}{10} \times 5 = .10 \times 5 = .50 \]

c. \[ P(12 \leq x \leq 18) = \frac{1}{10} \times 6 = .10 \times 6 = .60 \]

d. \[ E(x) = \frac{10 + 20}{2} = 15 \]

e. \[ Var(x) = \frac{(20 - 10)^2}{12} = 8.33 \]
4. a. 

\[
f(x) = \begin{cases} 
1.5 & \text{for } x < .25 \\
1.0 & \text{for } .25 < x < .75 \\
.5 & \text{for } x > .75 \\
0 & \text{elsewhere}
\end{cases}
\]

b. \( P(.25 < x < .75) = 1 \cdot (.50) = .50 \)

c. \( P(x \leq .30) = 1 \cdot (.30) = .30 \)

d. \( P(x > .60) = 1 \cdot (.40) = .40 \)

e./f. Answers will vary.

6. a. For a uniform probability density function 

\[
f(x) = \begin{cases} 
\frac{1}{b-a} & \text{for } a < x < b \\
0 & \text{elsewhere}
\end{cases}
\]

Thus, \( \frac{1}{b-a} = .00625 \).

Solving for \( b-a \), we have \( b-a = \frac{1}{.00625} = 160 \)

In a uniform probability distribution, \( \frac{1}{2} \) of this interval is below the mean and \( \frac{1}{2} \) of this interval is above the mean. Thus,

\[
a = 136 - \frac{1}{2}(160) = 56 \quad \text{and} \quad b = 136 + \frac{1}{2}(160) = 216
\]

b. \( P(100 < x < 200) = (200 - 100)(.00625) = .6250 \)

c. \( P(x \geq 150) = (216 - 150)(.00625) = .4125 \)

d. \( P(x \leq 80) = (80 - 56)(.00625) = .1500 \)

8. 

\[
\sigma = 10
\]

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
10. These probabilities can be obtained using Excel’s NORM.S.DIST function or the standard normal probability table in the text.
   a. \( P(z \leq 1.5) = 0.9332 = \text{NORM.S.DIST}(1.5, \text{TRUE}) \)
   b. \( P(z \leq 1.0) = 0.8413 = \text{NORM.S.DIST}(1.0, \text{TRUE}) \)
   c. \( P(1 \leq z \leq 1.5) = P(z \leq 1.5) - P(z < 1) = 0.9332 - 0.8413 = 0.0919 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(1.5, \text{TRUE}) - \text{NORM.S.DIST}(1.0, \text{TRUE}) = 0.0918
   \]
   d. \( P(0 < z < 2.5) = P(z < 2.5) - P(z \leq 0) = 0.9938 - 0.5000 = 0.4938 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(2.5, \text{TRUE}) - \text{NORM.S.DIST}(0, \text{TRUE})
   \]
12. These probabilities can be obtained using Excel’s NORM.S.DIST function or the standard normal probability table in the text.
   a. \( P(0 \leq z \leq 0.83) = P(z \leq 0.83) - P(z \leq 0) = 0.7967 - 0.5000 = 0.2967 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(0.83, \text{TRUE}) - \text{NORM.S.DIST}(0, \text{TRUE})
   \]
   b. \( P(-1.57 \leq z \leq 0) = P(z \leq 0) - P(z < -1.57) = 0.5000 - 0.0582 = 0.4418 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(0, \text{TRUE}) - \text{NORM.S.DIST}(-1.57, \text{TRUE})
   \]
   c. \( P(z > 0.44) = 1 - P(z \leq 0.44) = 1 - 0.6700 = 0.3300 = 1 - \text{NORM.S.DIST}(0.44, \text{TRUE}) \)
   d. \( P(z \geq -0.23) = 1 - P(z < -0.23) = 1 - 0.4090 = 0.5910 = 1 - \text{NORM.S.DIST}(-0.23, \text{TRUE}) \)
   e. \( P(z < 1.20) = 0.8849 = \text{NORM.S.DIST}(1.2, \text{TRUE}) \)
   f. \( P(z \leq -0.71) = 0.2389 = \text{NORM.S.DIST}(-0.71, \text{TRUE}) \)
13. These probabilities can be obtained using Excel’s NORM.S.DIST function or the standard normal probability table in the text.
   a. \( P(-1.98 \leq z \leq 0.49) = P(z \leq 0.49) - P(z < -1.98) = 0.6879 - 0.0239 = 0.6640 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(0.49, \text{TRUE}) - \text{NORM.S.DIST}(-1.98, \text{TRUE}) = 0.6641
   \]
   b. \( P(0.52 \leq z \leq 1.22) = P(z \leq 1.22) - P(z < 0.52) = 0.8888 - 0.6985 = 0.1903 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(1.22, \text{TRUE}) - \text{NORM.S.DIST}(0.52, \text{TRUE})
   \]
   c. \( P(-1.75 \leq z \leq -1.04) = P(z \leq -1.04) - P(z < -1.75) = 0.1492 - 0.0401 = 0.1091 \)
   \[
   \text{OR} = \text{NORM.S.DIST}(-1.04, \text{TRUE}) - \text{NORM.S.DIST}(-1.75, \text{TRUE})
   \]
14. These \( z \) values can be obtained using Excel’s NORM.S.INV function or by using the standard normal probability table in the text.
   a. The \( z \) value corresponding to a cumulative probability of 0.9750 is \( z = 1.96 \) = NORM.S.INV(0.975)
   b. Since the area to the left of \( z=0 \) is 0.5, the \( z \) value here also corresponds to a cumulative probability of 0.9750: \( z = 1.96 \) NORM.S.INV(0.975)
   c. The \( z \) value corresponding to a cumulative probability of 0.7291 is \( z = 0.51 \) = NORM.S.INV(0.7291)
   d. Area to the left of \( z \) is 1 - 0.1314 = 0.8686. So \( z = 1.12 \) = NORM.S.INV(0.8686)
e. The $z$ value corresponding to a cumulative probability of .6700 is $z = .44$. \text{NORM.S.INV(.67)}

f. The area to the left of $z$ is .6700. So $z = .44$. \text{NORM.S.INV(.67)}

15. These $z$ values can be obtained using Excel’s NORM.S.INV function or by using the standard normal probability table in the text.

a. The $z$ value corresponding to a cumulative probability of .2119 is $z = -.80$. \text{NORM.S.INV(.2119)}

b. Compute .9030/2 = .4515; $z$ corresponds to a cumulative probability of $.5000 + .4515 = .9515$. So $z = 1.66$. \text{NORM.S.INV(.9515)}

c. Compute .2052/2 = .1026; $z$ corresponds to a cumulative probability of $.5000 + .1026 = .6026$. So $z = .26$. \text{NORM.S.INV(.6026)}

d. The $z$ value corresponding to a cumulative probability of .9948 is $z = 2.56$. \text{NORM.S.INV(.9948)}

e. The area to the left of $z$ is 1 - .6915 = .3085. So $z = - .50$. \text{NORM.S.INV(.3085)}

16. These $z$ values can be obtained using Excel’s NORM.S.INV function or the standard normal probability table in the text.

a. The area to the left of $z$ is 1 - .0100 = .9900. The $z$ value in the table with a cumulative probability closest to .9900 is $z = 2.33$. \text{NORM.S.INV(.99)}

b. The area to the left of $z$ is .9750. So $z = 1.96$. \text{NORM.S.INV(.975)}

c. The area to the left of $z$ is .9500. Since .9500 is exactly halfway between .9495 ($z = 1.64$) and .9505($z = 1.65$), we select $z = 1.645$. However, $z = 1.64$ or $z = 1.65$ are also acceptable answers. \text{NORM.S.INV(.95)}

d. The area to the left of $z$ is .9000. So $z = 1.28$ is the closest $z$ value. \text{NORM.S.INV(.9)}

18. $\mu = 14.4$ and $\sigma = 4.4$

a. At $x = 20$, $z = \frac{20 - 14.4}{4.4} = 1.27$

$P(z \leq 1.27) = .8980$

$P(x \geq 20) = P(z \geq 1.27) = 1 - P(z \leq 1.27) = 1 - .8980 = .1020$

Using Excel: 1 - NORM.DIST(20,14.4,4.4,TRUE) = .1016

b. At $x = 10$, $z = \frac{10 - 14.4}{4.4} = -1.00$

$P(z \leq -1.00) = .1587$

So, $P(x \leq 10) = .1587$

Using Excel: 1 - NORM.DIST(10,14.4,4.4,TRUE) = .1587

c. A $z$-value of 1.28 cuts off an area of approximately 10% in the upper tail.
A return of 20.03% or higher will put a domestic stock fund in the top 10%

Using Excel: 1-NORM.INV(.9,14.4,4.4) = 20.0388 or 20.04%

20. a. United States: \( \mu = 3.73 \quad \sigma = .25 \)

At \( x = 3.50 \), \( z = \frac{3.5 - 3.73}{.25} = -.92 \)

\( P(z < -.92) = .1788 \)

So, \( P(x < 3.50) = .1788 \)

Using Excel: NORM.DIST(3.5,3.73,.25,TRUE) = .1788

b. Russia: \( \mu = 3.40 \quad \sigma = .20 \)

At \( x = 3.50 \), \( z = \frac{3.50 - 3.40}{.20} = .50 \)

\( P(z < .50) = .6915 \)

So, \( P(x < 3.50) = .6915 \)

Using Excel: NORM.DIST(3.5,3.40,.20,TRUE) = .6915

69.15% of the gas stations in Russia charge less than $3.50 per gallon.

c. Use mean and standard deviation for Russia.

At \( x = 3.73 \), \( z = \frac{3.73 - 3.40}{.20} = 1.65 \)

\( P(z > 1.65) = 1 - P(z \leq 1.65) = 1 - .9505 = .0495 \)

\( P(x > 3.73) = .0495 \)

Using Excel: 1-NORM.DIST(3.73,3.40,.20,TRUE) = .0495

The probability that a randomly selected gas station in Russia charges more than the mean price in the United States is .0495. Stated another way, only 4.95% of the gas stations in Russia charge more than the average price in the United States.

22. Use \( \mu = 8.35 \) and \( \sigma = 2.5 \)

a. We want to find \( P(5 \leq x \leq 10) \)

At \( x = 10 \),

\[
    z = \frac{x - \mu}{\sigma} = \frac{10 - 8.35}{2.5} = .66
\]

At \( x = 5 \),
z = \frac{x - \mu}{\sigma} = \frac{5 - 8.35}{2.5} = -1.34

P(5 \leq x \leq 10) = P(-1.34 \leq z \leq .66) = P(z \leq .66) - P(z \leq -1.34)
= .7454 - .0901
= .6553

Using Excel: NORM.DIST(10,8.35,2.5,TRUE) – NORM.DIST(5,8.35,2.5,TRUE) = .6553

The probability of a household viewing television between 5 and 10 hours a day is .6553.

b. Find the z-value that cuts off an area of .03 in the upper tail. Using a cumulative probability of 1 -.03 = .97, z = 1.88 provides an area of .03 in the upper tail of the normal distribution.

x = \mu + z \sigma = 8.35 + 1.88(2.5) = 13.05 hours

Using Excel: NORM.INV(.97,8.35,2.5) = 13.0520

A household must view slightly over 13 hours of television a day to be in the top 3% of television viewing households.

c. At x = 3, z = \frac{x - \mu}{\sigma} = \frac{3 - 8.35}{2.5} = -2.14

P(x>3) = P(z > -2.14) = 1 - P(z \leq -2.14) = 1 - .0162 = .9838

Using Excel: 1-NORM.DIST(3,8.35,2.5) = .9838

The probability a household views more than 3 hours of television a day is .9838.

24. \mu = 749 and \sigma = 225

a. z = \frac{x - \mu}{\sigma} = \frac{400 - 749}{225} = -1.55

P(x < 400) = P(z < -1.55) = .0606

The probability that expenses will be less than $400 is .0606.

Using Excel: NORM.DIST(400,749,225,TRUE) = .0604

b. z = \frac{x - \mu}{\sigma} = \frac{800 - 749}{225} = .23

P(x \geq 800) = P(z \geq .23) = 1 - P(z \leq .23) = 1 - .5910 = .4090

The probability that expenses will be $800 or more is .4090.

Using Excel: 1-NORM.DIST(800,749,225,TRUE) = .4103

c. For x = 1000, z = \frac{x - \mu}{\sigma} = \frac{1000 - 749}{225} = 1.12

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
For \( x = 500, \ z = \frac{x - \mu}{\sigma} = \frac{500 - 749}{225} = -1.11 \)

\[ P(500 \leq x \leq 1000) = P(-1.11 \leq z \leq 1.12) = P(z \leq 1.12) - P(z \leq -1.11) = .8686 - .1335 = .7351 \]

The probability that expenses will be between $500 and $1000 is .7351.

Using Excel: NORM.DIST(1000,749,225,TRUE) – NORM.DIST(500,749,225,TRUE) = .7335

d. The upper 5%, or area = 1 - .05 = .95 occurs for \( z = 1.645 \)

\[ x = \mu + z\sigma = 749 + 1.645(225) = $1119 \]

The 5% most expensive travel plans will be slightly more than $1119 or higher.

Using Excel: NORM.INV(.95,749,225) = 1119.0921

26. \( \mu = 8 \)

a. \( P(x \leq 6) = 1 - e^{-6/8} = 1 - .4724 = .5276 \)

Using Excel: EXPON.DIST(6,1/8,TRUE) = .5276

b. \( P(x \leq 4) = 1 - e^{-4/8} = 1 - .6065 = .3935 \)

Using Excel: EXPON.DIST(4,1/8,TRUE) = .3935

c. \( P(x \geq 6) = 1 - P(x \leq 6) = 1 - .5276 = .4724 \)

Using Excel: =1-EXPON.DIST(6,1/8,TRUE) = .4724

d. \( P(4 \leq x \leq 6) = P(x \leq 6) - P(x \leq 4) = .5276 - .3935 = .1342 \)

Using Excel: EXPON.DIST(6,1/8,TRUE)-EXPON.DIST(4,1/8,TRUE) = .1342

27. \( \mu = 3 \)

a. \( P(x \leq x_0) = 1 - e^{-x_0/3} \)

b. \( P(x \leq 2) = 1 - e^{-2/3} = 1 - .5134 = .4866 \)

Using Excel: EXPON.DIST(2,1/3,TRUE) = .4866

c. \( P(x \geq 3) = 1 - P(x \leq 3) = 1 - (1 - e^{-3/3}) = e^{-1} = .3679 \)

Using Excel: =1-EXPON.DIST(3,1/3,TRUE) = .3679

d. \( P(x \leq 5) = 1 - e^{-5/3} = 1 - .1889 = .8111 \)

Using Excel: EXPON.DIST(5,1/3,TRUE) = .8111

e. \( P(2 \leq x \leq 5) = P(x \leq 5) - P(x \leq 2) = .8111 - .4866 = .3245 \)

Using Excel: EXPON.DIST(5,1/3,TRUE)-EXPON.DIST(2,1/3,TRUE) = .3245
28. a. With $\mu = 20$, \( f(x) = \frac{1}{20} e^{-x/20} \)

b. \( P(x \leq 15) = 1 - e^{-x/\mu} = 1 - e^{-15/20} = .5276 = \text{EXPON.DIST}(15,1/20,\text{TRUE}) \)

c. \( P(x > 20) = 1 - P(x \leq 20) \)
\[ = 1 - (1 - e^{-20/20}) = e^{-1} = .3679 = 1 - \text{EXPON.DIST}(20,1/20,\text{TRUE}) \]

d. With $\mu = 7$, \( f(x) = \frac{1}{7} e^{-x/7} \)
\( P(x \leq 5) = 1 - e^{-x/\mu} = 1 - e^{-5/7} = .5105 = \text{EXPON.DIST}(5,1/7,\text{TRUE}) \)

29. a. \[
\begin{align*}
\text{f(x)} & \\
\mu = 12 \\
\text{b. } P(x \leq 12) & = 1 - e^{-12/12} = 1 - .3679 = .6321 \\
\text{Using Excel: } \text{EXPON.DIST}(12,1/12,\text{TRUE}) & = .6321 \\
\text{c. } P(x \leq 6) & = 1 - e^{-6/12} = 1 - .6065 = .3935 \\
\text{Using Excel: } \text{EXPON.DIST}(6,1/12,\text{TRUE}) & = .3935 \\
\text{d. } P(x \geq 30) & = 1 - P(x < 30) \\
& = 1 - (1 - e^{-30/12}) \\
& = .0821 \\
\text{Using Excel: } =1-\text{EXPON.DIST}(30,1/12,\text{TRUE}) & = .0821 \\
\end{align*}
\]

30. $\mu = 2$

a. \( f(x) = \frac{1}{\mu} e^{-x/\mu} = \frac{1}{2} e^{-x/2} \) for \( x \geq 0 \)
Continuous Probability Distributions

\[ P(x \leq x_0) = 1 - e^{-x_0/\mu} \]

\[ P(x \leq 1) = 1 - e^{-1/2} = 1 - e^{-5} = 1 - .6065 = .3935 = \text{EXPON.DIST}(1, 1/2, \text{TRUE}) \]

b. \[ P(x \leq 2) = 1 - e^{-2/2} = 1 - e^{-1.0} = 1 -.3679 = .6321 \]

\[ P(1 \leq x \leq 2) = P(x \leq 2) - P(x \leq 1) = .6321 - .3935 = .2386 = \text{EXPON.DIST}(2, 1/2, \text{TRUE}) - \text{EXPON.DIST}(1, 1/2, \text{TRUE}) = .2387 \]

c. For this customer, the cable service repair would have to take longer than 4 hours.

\[ P(x > 4) = 1 - P(x \leq 4) = 1 - (1 - e^{-4/2}) = e^{-2.0} = .1353 = 1 - \text{EXPON.DIST}(4, 1/2, \text{TRUE}) \]

32. a. Because the number of calls per hour follows a Poisson distribution, the time between calls follows an exponential distribution. So,

for a mean of 1.6 calls per hour, the mean time between calls is

\[ \mu = \frac{60 \text{ minutes/hour}}{1.6 \text{ calls/hour}} = 37.5 \text{ minutes per call} \]

b. The exponential probability density function is

\[ f(x) = \begin{cases} \frac{1}{37.5} e^{-x/37.5} & \text{for } x \geq 0 \\ \end{cases} \]

where \( x \) is the minutes between 911 calls.

c. Using time in minutes,

\[ P(x < 60) = 1 - e^{-60/37.5} = 1 -.2019 = .7981 = \text{EXPON.DIST}(60, 1/37.5, \text{TRUE}) \]

d. \[ P(x \geq 30) = 1 - P(x < 30) = 1 - (1 - e^{-30/37.5}) = 1 -.5507 = .4493 \]

\[ = 1 - \text{EXPON.DIST}(30, 1/37.5, \text{TRUE}) \]

e. \[ P(5 \leq x \leq 20) = (1 - e^{-20/37.5}) - (1 - e^{-5/37.5}) = .4134 - .1248 = .2885 \]

\[ \text{EXPON.DIST}(20, 1/37.5, \text{TRUE}) - \text{EXPON.DIST}(5, 1/37.5, \text{TRUE}) = .2885 \]

34. \( \mu = 19000 \) and \( \sigma = 2100 \)

a. Find the \( z \) value that cuts off an area of .10 in the lower tail.

From the standard normal table \( z = -1.28 \). Solve for \( x \),

\[ z = -1.28 = \frac{x - 19,000}{2100} \]

\[ x = 19,000 - 1.28(2100) = 16,312 \]

10% of athletic scholarships are valued at $16,312 or less.

Using Excel: \( \text{NORM.INV(.90, 19000, 2100)} = 16,308.74 \)
Chapter 6

b. 
\[ z = \frac{x - \mu}{\sigma} = \frac{22,000 - 19,000}{2100} = 1.43 \]

\[ P(x \geq 22,000) = P(z \geq 1.43) = 1 - P(z \leq 1.43) = 1 - .9236 = .0764 \]

7.64% of athletic scholarships are valued at $22,000 or more.

Using Excel: NORM.DIST(22000,19000,2100,TRUE) = .0766

c. Find the \( z \) value that cuts off an area of .03 in the upper tail: \( z = 1.88 \). Solve for \( x \),

\[ z = 1.88 = \frac{x - 19,000}{2100} \]

\[ x = 19,000 + 1.88(2100) = 22,948 \]

3% of athletic scholarships are valued at $22,948 or more.

Using Excel: NORM.INV(.97,19000,2100) = 22,949.6666

36. \( \mu = 658 \)

a. \( z = -1.88 \) cuts off .03 in the lower tail

So,

\[ z = -1.88 = \frac{610 - 658}{\sigma} \]

\[ \sigma = \frac{610 - 658}{-1.88} = 25.5319 \]

Using EXCEL: NORM.S.INV(.03), \( z = -1.8807936 \), solving for \( \sigma \) without rounding, gives 25.5211

b. At 700, \( z = \frac{x - \mu}{\sigma} = \frac{700 - 658}{25.5319} = 1.65 \)

At 600, \( z = \frac{x - \mu}{\sigma} = \frac{600 - 659}{25.5319} = -2.31 \)

\[ P(600 < x < 700) = P(-2.31 < z < 1.65) = P(z < 1.65) - P(z < -2.31) = .9505 - .0104 = .9401 \]

Using Excel: NORM.DIST(700,658,25.5211,TRUE) - NORM.DIST(600,658,25.5211,TRUE) = .9386

c. \( z = 1.88 \) cuts off approximately .03 in the upper tail

\[ x = 658 + 1.88(25.5319) = 706. \]

Using Excel: NORM.INV(.97,658,25.5211) = 706

On the busiest 3% of days 706 or more people show up at the pawnshop.
38. a. At $x = 200$

$$z = \frac{200 - 150}{25} = 2$$

$$P(x > 200) = P(z > 2) = 1 - P(z \leq 2) = 1 - .9772 = .0228 = 1 - \text{NORM.DIST}(200,150,25,\text{TRUE})$$

b. Expected Profit = Expected Revenue - Expected Cost

$$= 200 - 150 = 50$$

40. $\mu = 450$ and $\sigma = 100$

a. At 400,

$$z = \frac{400 - 450}{100} = -.50$$

Area to left is .3085

At 500,

$$z = \frac{500 - 450}{100} = .50$$

Area to left is .6915

$$P(400 \leq x \leq 500) = P(-.5 \leq z \leq .5) = P(z \leq .5) - P(z \leq -.5) = .6915 - .3085 = .3830$$

Using Excel: $\text{NORM.DIST}(500,450,100,\text{TRUE}) - \text{NORM.DIST}(400,450,100,\text{TRUE}) = .3829$

38.3% will score between 400 and 500.

b. At 630,

$$z = \frac{630 - 450}{100} = 1.80$$

Probability of worse than 630 = $P(x < 630) = P(z < 1.8) = .9641$

Using Excel: $\text{NORM.DIST}(630,450,100,\text{TRUE}) = .9641$

Probability of better than 630 = $P(x > 630) = P(z > 1.8) = 1 - P(z \leq 1.8) = 1 - .9641 = .0359$

Using Excel: $1 - \text{NORM.DIST}(630,450,100,\text{TRUE}) = .0359$

96.41% do worse and 3.59% do better.

c. At 480,

$$z = \frac{480 - 450}{100} = .30$$

Area to left is .6179

Probability of admittance = $P(x \geq 480) = P(z \geq .3) = 1 - P(z \leq .3) = 1 - .6179 = .3821$

Using Excel: $1 - \text{NORM.DIST}(480,450,100,\text{TRUE}) = .3821$

38.21% are acceptable.
42. \( \sigma = .6 \)

At 2% \( z \approx -2.05 \)

\[
\begin{align*}
\frac{x - \mu}{\sigma} &= -2.05 \\
\Rightarrow x &= 18 - \mu \\
\mu &= 18 - 2.05(0.6) = 19.23 \text{ oz.}
\end{align*}
\]

The mean filling weight must be 19.23 oz.

44. a. Mean time between arrivals = 1/7 minutes

b. \( f(x) = 7e^{-7x} \)

c. \( P(\text{no one in 1 minute}) = P(\text{greater than 1 minute between arrivals}) = 1 - \text{EXPON.DIST}(1,1/(1/7),TRUE) = .0009 \)

Note using Poisson via Excel gives: POISSON.DIST(0,7,TRUE) = .0009

d. 12 seconds is .2 minutes, or 1/5 of a minute, therefore poisson mean = 7/5 per 12 seconds = 1.4

Using exponential, \( P(\text{no one in 12 seconds}) = P(\text{greater than 12 seconds between arrivals}) = 1 - \text{EXPON.DIST}(0.2,1/(1/7),TRUE) = .2466 \)

Note using Poisson via Excel gives: POISSON.DIST(0,7/5,TRUE) = .2466

46. a. \( \frac{1}{\mu} = 0.5 \) therefore \( \mu = 2 \) minutes = mean time between telephone calls

b. Note: 30 seconds = .5 minutes

\( P(x \leq .5) = 1 - e^{-.5/2} = 1 - .7788 = .2212 = \text{EXPON.DIST}(.5,1/2,TRUE) \)

c. \( P(x \leq 1) = 1 - e^{-1/2} = 1 - .6065 = .3935 = \text{EXPON.DIST}(1,1/2,TRUE) \)

d. \( P(x \geq 5) = 1 - P(x < 5) = 1 - (1 - e^{-5/2}) = .0821 = \text{EXPON.DIST}(5,1/2,TRUE) \)
Chapter 7
Sampling and Sampling Distributions

Solutions:

1. a. AB, AC, AD, AE, BC, BD, BE, CD, CE, DE
   b. With 10 samples, each has a 1/10 probability.
   c. B and D because the two smallest random numbers are .0476 and .0957.

2. The 4 smallest random numbers are .0341, .0729, .0936, and .1449. So elements 2, 3, 5, and 10 are the simple random sample.

3. The simple random sample consists of New York, Detroit, Oakland, Boston, and Kansas City.

4. Step 1: Generate a random number using the RAND() function for each of the 10 golfers.
   Step 2: Sort the list of golfers with respect to the random numbers. The first 3 golfers in the sorted list make up the simple random sample. Answers will vary with every regeneration of random numbers.

6. a. Finite population. A frame could be constructed obtaining a list of licensed drivers from the New York state driver’s license bureau.
   b. Sampling from an infinite population. The sample is taken from the production line producing boxes of cereal.
   c. Sampling from an infinite population. The sample is taken from the ongoing arrivals to the Golden Gate Bridge.
   d. Finite population. A frame could be constructed by obtaining a listing of students enrolled in the course from the professor.
   e. Sampling from an infinite population. The sample is taken from the ongoing orders being processed by the mail-order firm.

7. a. \( \bar{x} = \frac{\sum x_i}{n} = \frac{54}{6} = 9 \)
   b. \( s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \)
      \( \sum (x_i - \bar{x})^2 = (-4)^2 + (-1)^2 + 1^2 + (-2)^2 + 1^2 + 5^2 = 48 \)
      \( s = \sqrt{\frac{48}{6-1}} = 3.1 \)
8. a. $\hat{p} = \frac{75}{150} = .50$

   b. $\hat{p} = \frac{55}{150} = .3667$

9. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{465}{5} = 93$

   b.

   \[
   \begin{array}{ccc}
   x_i & (x_i - \bar{x}) & (x_i - \bar{x})^2 \\
   94 & +1 & 1 \\
   100 & +7 & 49 \\
   85 & -8 & 64 \\
   94 & +1 & 1 \\
   92 & -1 & 1 \\
   \text{Totals} & 465 & 0 & 116 \\
   \end{array}
   \]

   \[s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{116}{4}} = 5.39\]

10. a. Two of the 40 stocks in the sample received a 5 Star rating.

    \[\hat{p} = \frac{2}{40} = .05\]

   b. Seventeen of the 40 stocks in the sample are rated Above Average with respect to risk.

    \[\hat{p} = \frac{17}{40} = .425\]

   c. There are eight stocks in the sample that are rated 1 Star or 2 Star.

    \[\hat{p} = \frac{8}{40} = .20\]

12. a. The sampled population is U. S. adults that are 50 years of age or older.

   b. We would use the sample proportion for the estimate of the population proportion.

    \[\hat{p} = \frac{350}{426} = .8216\]

   c. The sample proportion for this issue is .74 and the sample size is 426.

    The number of respondents citing education as “very important” is (.74)*426 = 315.

   d. We would use the sample proportion for the estimate of the population proportion.

    \[\hat{p} = \frac{354}{426} = .8310\]

   e. The inferences in parts (b) and (d) are being made about the population of U.S. adults who are age 50 or older. So, the population of U.S. adults who are age 50 or older is the target population. The target population is the same as the sampled population. If the sampled population was
restricted to members of AARP who were 50 years of age or older, the sampled population would not be the same as the target population. The inferences made in parts (b) and (d) would only be valid if the population of AARP members age 50 or older was representative of the U.S. population of adults age 50 and over.

14. a. Use the data disk accompanying the book and the EAI file. Generate a random number for each manager and select managers associated with the 50 smallest random numbers as the sample. Answers will vary with every regeneration of random numbers.

b. Use Excel's AVERAGE function to compute the mean for the sample.

c. Use Excel's STDEV.S function to compute the sample standard deviation.

d. Use the sample proportion as a point estimate of the population proportion.

15. a. The sampling distribution is normal with

\[ E(\bar{x}) = \mu = 200 \]

\[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 50 / \sqrt{100} = 5 \]

For \( \pm 5, 195 \leq \bar{x} \leq 205 \)

Using Standard Normal Probability Table:

At \( \bar{x} = 205 \), \( z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{5}{5} = 1 \ P(z \leq 1) = .8413 \)

At \( \bar{x} = 195 \), \( z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{-5}{5} = -1 \ P(z < -1) = .1587 \)

\[ P(195 \leq \bar{x} \leq 205) = .8413 - .1587 = .6826 \]

Using Excel: \( \text{NORM.DIST}(205,200,50/\sqrt{100},\text{TRUE}) - \text{NORM.DIST}(195,200,50/\sqrt{100},\text{TRUE}) = .6827 \)

b. For \( \pm 10, 190 \leq \bar{x} \leq 210 \)

Using Standard Normal Probability Table:

At \( \bar{x} = 210 \), \( z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{10}{5} = 2 \ P(z \leq 2) = .9772 \)

At \( \bar{x} = 190 \), \( z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{-10}{5} = -2 \ P(z < -2) = .0228 \)

\[ P(190 \leq \bar{x} \leq 210) = .9772 - .0228 = .9544 \]

Using Excel: \( \text{NORM.DIST}(210,200,50/\sqrt{100},\text{TRUE}) - \text{NORM.DIST}(190,200,50/\sqrt{100},\text{TRUE}) = .9545 \)
16. \( \sigma_\bar{x} = \sigma / \sqrt{n} \)

\( \sigma_\bar{x} = 25 / \sqrt{50} = 3.54 \)

For sample size = 100, \( \sigma_\bar{x} = 25 / \sqrt{100} = 2.50 \)

For sample size = 150, \( \sigma_\bar{x} = 25 / \sqrt{150} = 2.04 \)

For sample size = 200, \( \sigma_\bar{x} = 25 / \sqrt{200} = 1.77 \)

The standard error of the mean decreases as the sample size increases.

18. a. The normal distribution for \( \bar{x} \) is based on the Central Limit Theorem.

b. For \( n = 120 \), \( E(\bar{x}) \) remains \$51,800 and the sampling distribution of \( \bar{x} \) can still be approximated by a normal distribution. However, \( \sigma_\bar{x} \) is reduced to \( 4000 / \sqrt{120} = 365.15 \).

c. As the sample size is increased, the standard error of the mean, \( \sigma_\bar{x} \), is reduced. This appears logical from the point of view that larger samples should tend to provide sample means that are closer to the population mean. Thus, the variability in the sample mean, measured in terms of \( \sigma_\bar{x} \), should decrease as the sample size is increased.

19. a. With a sample of size 60 \( \sigma_\bar{x} = \frac{4000}{\sqrt{60}} = 516.40 \)

At \( \bar{x} = 52,300 \), \( z = \frac{52,300 - 51,800}{516.40} = .97 \)

\( P(\bar{x} \leq 52,300) = P(z \leq .97) = .8340 \)

At \( \bar{x} = 51,300 \), \( z = \frac{51,300 - 51,800}{516.40} = -.97 \)
\[ P(\bar{x} < 51,300) = P(z < -0.97) = 0.1660 \]
\[ P(51,300 \leq \bar{x} \leq 52,300) = 0.8340 - 0.1660 = 0.6680 \]

Using Excel: \( \text{NORM.DIST}(52300,51800,4000/\sqrt{60},\text{TRUE}) - \text{NORM.DIST}(51300,51800,4000/\sqrt{60},\text{TRUE}) = 0.6671 \)

20. a. Normal distribution, \( E(\bar{x}) = \mu = 17.5 \)

\[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 4 / \sqrt{50} = 0.5657 \]

b. Within 1 week means \( 16.5 \leq \bar{x} \leq 18.5 \)

At \( \bar{x} = 18.5, \quad z = \frac{18.5 - 17.5}{0.5657} = 1.77. \quad P(z \leq 1.77) = 0.9616 \)

At \( \bar{x} = 16.5, \quad z = -1.77. \quad P(z < -1.77) = 0.0384 \)

So \( P(16.5 \leq \bar{x} \leq 18.5) = 0.9616 - 0.0384 = 0.9232 \)

Using Excel: \( \text{NORM.DIST}(18,5,17.5,4/\sqrt{50},\text{TRUE}) - \text{NORM.DIST}(16.5,17.5,4/\sqrt{50},\text{TRUE}) = 0.9229 \)

c. Within 1/2 week means \( 17.0 \leq \bar{x} \leq 18.0 \)

At \( \bar{x} = 18.0, \quad z = \frac{18.0 - 17.5}{0.5657} = 0.88. \quad P(z \leq 0.88) = 0.8106 \)

At \( \bar{x} = 17.0, \quad z = -0.88. \quad P(z < -0.88) = 0.1894 \)

\[ P(17.0 \leq \bar{x} \leq 18.0) = 0.8106 - 0.1894 = 0.6212 \]

Using Excel: \( \text{NORM.DIST}(18,17.5,4/\sqrt{50},\text{TRUE}) - \text{NORM.DIST}(16,17.5,4/\sqrt{50},\text{TRUE}) = 0.6232 \)

22. a. \[ z = \frac{\bar{x} - 16642}{\sigma / \sqrt{n}} \]

Within \( \pm 200 \) means \( \bar{x} - 16642 \) must be between -200 and +200.

The \( z \) value for \( \bar{x} - 16642 = -200 \) is the negative of the \( z \) value for \( \bar{x} - 16642 = 200 \). So we just show the computation of \( z \) for \( \bar{x} - 16642 = 200 \).

\[ n = 30 \quad z = \frac{200}{2400 / \sqrt{30}} = 0.46 \quad P(-0.46 \leq z \leq 0.46) = 0.6772 - 0.3228 = 0.3544 \]

Using Excel: \( \text{NORM.DIST}(16842,16642,2400/\sqrt{30},\text{TRUE}) - \text{NORM.DIST}(16442,16642,2400/\sqrt{30},\text{TRUE}) = 0.3519 \)

\[ n = 50 \quad z = \frac{200}{2400 / \sqrt{50}} = 0.59 \quad P(-0.59 \leq z \leq 0.59) = 0.7224 - 0.2776 = 0.4448 \]
Chapter 7

Using Excel: \( \text{NORM.DIST}(16842,16642,2400/\text{SQRT}(50),\text{TRUE})-\text{NORM.DIST}(16442,16642,2400/\text{SQRT}(50),\text{TRUE}) = .4443 \)

\( n = 100 \quad z = \frac{200}{2400/\sqrt{100}} = .83 \quad P(-.83 \leq z \leq .83) = .7967 - .2033 = .5934 \)

Using Excel: \( \text{NORM.DIST}(16842,16642,2400/\text{SQRT}(100),\text{TRUE})-\text{NORM.DIST}(16442,16642,2400/\text{SQRT}(100),\text{TRUE}) = .5953 \)

\( n = 400 \quad z = \frac{200}{2400/\sqrt{400}} = 1.67 \quad P(-1.67 \leq z \leq 1.67) = .9525 - .0475 = .9050 \)

Using Excel: \( \text{NORM.DIST}(16842,16642,2400/\text{SQRT}(400),\text{TRUE})-\text{NORM.DIST}(16442,16642,2400/\text{SQRT}(400),\text{TRUE}) = .9044 \)

b. A larger sample increases the probability that the sample mean will be within a specified distance of the population mean. In this instance, the probability of being within ±200 of \( \mu \) ranges from .3544 for a sample of size 30 to .9050 for a sample of size 400.

24. a. This is a graph of a normal distribution with \( E(\bar{x}) = \mu = 22 \) and

\[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 4 / \sqrt{30} = .7303 \]

b. Within 1 inch means 21 \( \leq \bar{x} \leq 23 \)

\[ z = \frac{23 - 22}{.7303} = 1.37 \quad z = \frac{21 - 22}{.7303} = -1.37 \]

\[ P(21 \leq \bar{x} \leq 23) = P(-1.37 \leq z \leq 1.37) = .9147 - .0853 = .8294 \]

The probability the sample mean will be within 1 inch of the population mean of 22 is .8294.

Using Excel: \( \text{NORM.DIST}(23,22,4/\text{SQRT}(30),\text{TRUE})-\text{NORM.DIST}(21,22,4/\text{SQRT}(30),\text{TRUE}) = .8291 \)

c. \[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 4 / \sqrt{45} = .5963 \]

Within 1 inch means 41 \( \leq \bar{x} \leq 43 \)

\[ z = \frac{43 - 42}{.5963} = 1.68 \quad z = \frac{41 - 42}{.5963} = -1.68 \]

\[ P(41 \leq \bar{x} \leq 43) = P(-1.68 \leq z \leq 1.68) = .9535 - .0465 = .9070 \]

The probability the sample mean will be within 1 inch of the population mean of 42 is .9070.

Using Excel: \( \text{NORM.DIST}(43,42,4/\text{SQRT}(45),\text{TRUE})-\text{NORM.DIST}(41,42,4/\text{SQRT}(45),\text{TRUE}) = .9065 \)

d. The probability of being within 1 inch is greater for New York in part (c) because the sample size is larger.
26. a. \( n / N = 40 / 4000 = .01 < .05 \); therefore, the finite population correction factor is not necessary.

b. With the finite population correction factor

\[
\sigma_x = \frac{n}{N-1} \sqrt{n} = \sqrt{\frac{4000 - 40}{4000 - 1} \frac{8.2}{40}} = 1.29
\]

Without the finite population correction factor

\[
\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{8.2}{\sqrt{40}} = 1.30
\]

Including the finite population correction factor provides only a slightly different value for \( \sigma_x \) than when the correction factor is not used.

c. Even though the population mean is not given, stating “within ±2” equates to “\( \bar{x} - \mu \)”.

\[
z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\pm 2}{8.2 / \sqrt{40}} = \pm 1.54 P(z \leq 1.54) = .9382
\]

\[
P(z < -1.54) = .0618
\]

Probability = .9382 - .0618 = .8764

Using Excel:

No value for the population mean is given, but we can use Excel’s NORM.S.DIST function here.

\[
NORM.S.DIST(2/(8.2/SQRT(40)),TRUE)-NORM.S.DIST(-2/(8.2/SQRT(40)),TRUE) = .8771
\]

28. a. \( E(\bar{p}) = p = .40 \)

\[
\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.40(.60)}{200}} = .0346
\]

Within ± .03 means \( .37 \leq \bar{p} \leq .43 \)

\[
z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.03}{.0346} = .87 P(z \leq .87) = .8078
\]

\[
P(z < .87) = .1922
\]

\[
P(.37 \leq \bar{p} \leq .43) = .8078 - .1922 = .6156
\]


b. \( z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.05}{.0346} = 1.44 P(z \leq 1.44) = .9251 \)

\[
P(z < 1.44) = .0749
\]
\( P(0.35 \leq \overline{p} \leq 0.45) = 0.9251 - 0.0749 = 0.8502 \)

Using Excel:\textsc{NORM.DIST}(0.45, 0.40, SQRT(0.4*0.6/200), TRUE) -
\textsc{NORM.DIST}(0.35, 0.40, SQRT(0.4*0.6/200), TRUE) = 0.8511

30. \( E(\overline{p}) = p = 0.30 \)

a. \( \sigma_{\overline{p}} = \sqrt{\frac{(0.30)(0.70)}{100}} = 0.0458 \)

Within ± 0.04 means 0.26 ≤ \( \overline{p} \) ≤ 0.34

\[ z = \frac{\overline{p} - p}{\sigma_{\overline{p}}} = \frac{0.04}{0.0458} = 0.87 \]
\[ P(z \leq 0.87) = 0.8078 \]

\( P(z < -0.87) = 0.1922 \)

\( P(0.26 \leq \overline{p} \leq 0.34) = 0.8078 - 0.1922 = 0.6156 \)

Using Excel:\textsc{NORM.DIST}(0.34, 0.30, SQRT(0.3*0.7/100), TRUE) -
\textsc{NORM.DIST}(0.26, 0.30, SQRT(0.3*0.7/100), TRUE) = 0.6173

b. \( \sigma_{\overline{p}} = \sqrt{\frac{(0.30)(0.70)}{200}} = 0.0324 \)

\[ z = \frac{\overline{p} - p}{\sigma_{\overline{p}}} = \frac{0.04}{0.0324} = 1.23 \]
\[ P(z \leq 1.23) = 0.8907 \]

\( P(z < -1.23) = 0.1093 \)

\( P(0.26 \leq \overline{p} \leq 0.34) = 0.8907 - 0.1093 = 0.7814 \)

Using Excel:\textsc{NORM.DIST}(0.34, 0.30, SQRT(0.3*0.7/200), TRUE) -
\textsc{NORM.DIST}(0.26, 0.30, SQRT(0.3*0.7/200), TRUE) = 0.7830

c. \( \sigma_{\overline{p}} = \sqrt{\frac{(0.30)(0.70)}{500}} = 0.0205 \)

\[ z = \frac{\overline{p} - p}{\sigma_{\overline{p}}} = \frac{0.04}{0.0205} = 1.95 \]
\[ P(z \leq 1.95) = 0.9744 \]

\( P(z < -1.95) = 0.0256 \)

\( P(0.26 \leq \overline{p} \leq 0.34) = 0.9744 - 0.0256 = 0.9488 \)

Using Excel:\textsc{NORM.DIST}(0.34, 0.30, SQRT(0.3*0.7/500), TRUE) -
\textsc{NORM.DIST}(0.26, 0.30, SQRT(0.3*0.7/500), TRUE) = 0.9490

d. \( \sigma_{\overline{p}} = \sqrt{\frac{(0.30)(0.70)}{1000}} = 0.0145 \)
\[ z = \frac{\bar{p} - p}{\sigma_p} = \frac{.04}{.0145} = 2.76 \quad P(z \leq 2.76) = .9971 \]

\[ P(z < -2.76) = .0029 \]

\[ P(.26 \leq \bar{p} \leq .34) = .9971 - .0029 = .9942 \]

Using Excel: \( \text{NORM.DIST}(.34,.30,\text{SQRT}(.3*\.7/1000),\text{TRUE}) - \text{NORM.DIST}(.26,.30,\text{SQRT}(.3*\.7/1000),\text{TRUE}) = .9942 \)

e. With a larger sample, there is a higher probability \( \bar{p} \) will be within \( \pm .04 \) of the population proportion \( p \).

31. a. The normal distribution is appropriate because \( np = 100(\.30) = 30 \) and \( n(1 - p) = 100(\.70) = 70 \) are both greater than 5.

b. \( P(.20 \leq \bar{p} \leq .40) = ? \)

\[ z = \frac{.40 - .30}{.0458} = 2.18 \quad P(z \leq 2.18) = .9854 \]

\[ P(z < -2.18) = .0146 \]

\[ P(.20 \leq \bar{p} \leq .40) = .9854 - .0146 = .9708 \]

Using Excel: \( \text{NORM.DIST}(.40,.30,\text{SQRT}(.3*\.7/100),\text{TRUE}) - \text{NORM.DIST}(.20,.30,\text{SQRT}(.3*\.7/100),\text{TRUE}) = .9709 \)

c. \( P(.25 \leq \bar{p} \leq .35) = ? \)

\[ z = \frac{.35 - .30}{.0458} = 1.09 \quad P(z \leq 1.09) = .8621 \]

\[ P(z < -1.09) = .1379 \]

\[ P(.25 \leq \bar{p} \leq .35) = .8621 - .1379 = .7242 \]

Using Excel: \( \text{NORM.DIST}(.35,.30,\text{SQRT}(.3*\.7/100),\text{TRUE}) - \text{NORM.DIST}(.25,.30,\text{SQRT}(.3*\.7/100),\text{TRUE}) = .7242 \)
NORM.DIST(.25,.30,SQRT(.3*.7/100),TRUE)=.7248

32. a. This is a graph of a normal distribution with a mean of $E(\bar{p}) = p = .55$ and

$$\sigma_\bar{p} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.55(1-.55)}{200}} = .0352$$

The normal distribution is appropriate because $np = 200(.55) = 110$ and $n(1-p) = 200(.45) = 90$ are both greater than 5.

b. Within $\pm .05$ means $.50 \leq \bar{p} \leq .60$

$$z = \frac{\bar{p} - p}{\sigma_\bar{p}} = \frac{.60 - .55}{.0352} = 1.42 \quad z = \frac{\bar{p} - p}{\sigma_\bar{p}} = \frac{.50 - .55}{.0352} = -1.42$$

$P(.50 \leq \bar{p} \leq .60) = P(-1.42 \leq z \leq 1.42) = .9222 - .0778 = .8444$

Using Excel: NORM.DIST(.60,.55,SQRT(.55*.45/200),TRUE) - NORM.DIST(.50,.55,SQRT(.55*.45/200),TRUE) = .8448

c. This is a graph of a normal distribution with a mean of $E(\bar{p}) = p = .45$ and

$$\sigma_\bar{p} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.45(1-.45)}{200}} = .0352$$

The normal distribution is appropriate because $np = 200(.45) = 90$ and $n(1-p) = 200(.55) = 110$ are both greater than 5.

d. $\sigma_\bar{p} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.45(1-.45)}{200}} = .0352$

Within $\pm .05$ means $.40 \leq \bar{p} \leq .50$

$$z = \frac{\bar{p} - p}{\sigma_\bar{p}} = \frac{.50 - .45}{.0352} = 1.42 \quad z = \frac{\bar{p} - p}{\sigma_\bar{p}} = \frac{.40 - .45}{.0352} = -1.42$$

$P(.40 \leq \bar{p} \leq .50) = P(-1.42 \leq z \leq 1.42) = .9222 - .0778 = .8444$

Using Excel: NORM.DIST(.50,.45,SQRT(.45*.55/200),TRUE) - NORM.DIST(.40,.45,SQRT(.45*.55/200),TRUE) = .8448

e. No, the probabilities are exactly the same. This is because $\sigma_\bar{p}$, the standard error, and the width of the interval are the same in both cases. Notice the formula for computing the standard error. It involves $p(1-p)$. So whenever $p(1-p)$ does not change, the standard error will be the same. In part (b), $p = .55$ and $1-p = .45$. In part (d), $p = .45$ and $1-p = .55$. 

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
f. For \( n = 400 \), \( \sigma_p = \sqrt{\frac{p(1 - p)}{n}} = \sqrt{\frac{.55(1 - .55)}{400}} = .0249 \)

Within \( \pm .05 \) means \(.50 \leq \bar{p} \leq .60\)

\[
\frac{\bar{p} - p}{\sigma_p} = \frac{.60 - .55}{.0249} = 2.01 \\
\frac{\bar{p} - p}{\sigma_p} = \frac{.50 - .55}{.0249} = -2.01
\]

\( P(.50 \leq \bar{p} \leq .60) = P(-2.01 \leq z \leq 2.01) = .9778 - .0222 = .9556 \)

Using Excel: \( \text{NORM.DIST(.60,.55,}\sqrt{.55*.45/400},\text{TRUE}) - \text{NORM.DIST(.50,.55,}\sqrt{.55*.45/400},\text{TRUE}) = .9556 \)

The probability is larger than in part (b). This is because the larger sample size has reduced the standard error from \(.0352\) to \(.0249\).

34. a. It is a normal distribution with \( E(\bar{p}) = p = .42 \) and

\[
\sigma_p = \sqrt{\frac{p(1 - p)}{n}} = \sqrt{\frac{.42(1 - .58)}{300}} = .0285
\]

The normal distribution is appropriate because \( np = 300(.42) = 126 \) and \( n(1 - p) = 300(.58) = 174 \) are both greater than 5.

b. \( z = \frac{\bar{p} - p}{\sigma_p} = \frac{.03}{.0285} = 1.05 P(z \leq 1.05) = .8531 \)

\( P(z < -1.05) = .1469 \)

\( P(.39 \leq \bar{p} \leq .45) = .8531 - .1469 = .7062 \)

Using Excel: \( \text{NORM.DIST(.45,.42,}\sqrt{.42*.58/300},\text{TRUE}) - \text{NORM.DIST(.39,.42,}\sqrt{.42*.58/300},\text{TRUE}) = .7076 \)

c. \( z = \frac{\bar{p} - p}{\sigma_p} = \frac{.05}{.0285} = 1.75 P(z \leq 1.75) = .9599 \)

\( P(z < -1.75) = .0401 \)

\( P(.37 \leq \bar{p} \leq .47) = .9599 - .0401 = .9198 \)

Using Excel: \( \text{NORM.DIST(.47,.42,}\sqrt{.42*.58/300},\text{TRUE}) - \text{NORM.DIST(.37,.42,}\sqrt{.42*.58/300},\text{TRUE}) = .9207 \)

d. The probabilities would increase. This is because the increase in the sample size makes the standard error, \( \sigma_p \), smaller.

36. a. \( E(\bar{p}) = p = .76 \)

\[
\sigma_p = \sqrt{\frac{p(1 - p)}{n}} = \sqrt{\frac{.76(1 - .76)}{400}} = .0214
\]
Chapter 7

Normal distribution because \( np = 400(.76) = 304 \) and \( n(1 - p) = 400(.24) = 96 \)

b. \( z = \frac{.79 - .76}{.0214} = 1.40 \) \( P(z \leq 1.40) = .9192 \)

\( P(z < -1.40) = .0808 \)

\( P(.73 \leq \bar{p} \leq .79) = P(-1.40 \leq z \leq 1.40) = .9192 - .0808 = .8384 \)

Using Excel: \( \text{NORM.DIST}(.79,.76,\sqrt{.76*.24/100},\text{TRUE}) - \text{NORM.DIST}(.73,.76,\sqrt{.76*.24/100},\text{TRUE})\) = .8399

c. \( \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.76(1-.76)}{750}} = .0156 \)

\( z = \frac{.79 - .76}{.0156} = 1.92 \) \( P(z \leq 1.92) = .9726 \)

\( P(z < -1.92) = .0274 \)

\( P(.73 \leq \bar{p} \leq .79) = P(-1.92 \leq z \leq 1.92) = .9726 - .0274 = .9452 \)

Using Excel: \( \text{NORM.DIST}(.79,.76,\sqrt{.76*.24/750},\text{TRUE}) - \text{NORM.DIST}(.73,.76,\sqrt{.76*.24/750},\text{TRUE})\) = .9456

38. a. 

The normal distribution for \( \bar{x} \) is based on the Central Limit Theorem. For \( n = 50 \), \( E(x) \) is 30, \( \sigma_{\bar{x}} \) is \( 6/\sqrt{50} = 0.8485 \), and the sampling distribution of \( \bar{x} \) can be approximated by a normal distribution.

b. 

\[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 6/\sqrt{50} = 0.8485 \]
The normal distribution for $\bar{x}$ is based on the Central Limit Theorem. For $n = 500000$, $E(\bar{x})$ remains 30 and the sampling distribution of $\bar{x}$ can be approximated by a normal distribution. However, now $\sigma_{\bar{x}} = 6/\sqrt{500000} = 0.0085$.

c. When the sample size is extremely large, the standard error of the sampling distribution of $\bar{x}$ becomes very small. This is logical because larger samples tend to provide sample means that are closer to the population mean. Thus, the variability in the sample mean, measured in terms of $\sigma_{\bar{x}}$, should decrease as the sample size is increased and should become very small when the sample size is extremely large.

40. a. $E(\bar{p}) = .37$ and $\sigma_{\bar{p}} = \sqrt{p(1-p)/n} = .0279$. $\bar{p}$ is also approximately normal because $np = 111 \geq 5$ and $n(1-p) = 189 \geq 5$.

b. At $\bar{p} = .32$, $z = \frac{.32 - .37}{.0279} = -1.79$

$P(\bar{p} \leq .32) = P(z \leq -1.79) = .0367$

At $\bar{p} = .42$, $z = \frac{.42 - .37}{.0279} = 1.79$

$P(\bar{p} < .42) = P(z < 1.79) = .9633$

$P(.42 \leq \bar{p} \leq .32) = .9633 - .0367 = .9266$

Using Excel: NORM.DIST(.42,.37,SQRT(.37*.63/300),TRUE) - NORM.DIST(.32,.37,SQRT(.37*.63/300),TRUE) = .9271

c. $E(\bar{p}) = .37$ and $\sigma_{\bar{p}} = \sqrt{p(1-p)/n} = .0028$. $\bar{p}$ is also approximately normal because $np = 11,100 \geq 5$ and $n(1-p) = 18,900 \geq 5$.

d. At $\bar{p} = .32$, $z = \frac{.32 - .37}{.0028} = -17.94$

$P(\bar{p} \leq .32) = P(z \leq -17.94) \approx .0000$
At $\bar{p} = .42$, $z = \frac{.42 - .37}{.0028} = 17.94$

$P(\bar{p} < .42) = P(z < 17.94) \approx 1.0000$

$P(.32 \leq \bar{p} \leq .42) \approx 1.0000 - .0000 = 1.0000$

Using Excel: $\text{NORM.DIST}(.42,.37,\text{SQRT}(.37*.63/30000),\text{TRUE}) - \text{NORM.DIST}(.32,.37,\text{SQRT}(.37*.63/30000),\text{TRUE}) = 1.0000$

e. The probability in part (d) is greater than the probability in part (b) because the larger sample size in part (d) results in a smaller standard error.

42. a. Sorting the list of companies and random numbers to identify the five companies associated with the five smallest random numbers provides the following sample.

<table>
<thead>
<tr>
<th>Company</th>
<th>Random Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMI Aerospace</td>
<td>.008012</td>
</tr>
<tr>
<td>Alpha &amp; Omega Semiconductor</td>
<td>.055369</td>
</tr>
<tr>
<td>Olympic Steel</td>
<td>.059279</td>
</tr>
<tr>
<td>Kimball International</td>
<td>.144127</td>
</tr>
<tr>
<td>International Shipholding</td>
<td>.227759</td>
</tr>
</tbody>
</table>

b. Step 1: Generate a new set of random numbers in column B. Step 2: Sort the random numbers and corresponding company names into ascending order and then select the companies associated with the five smallest random numbers. It is extremely unlikely that you will get the same companies as in part (a). Answers will vary with every regeneration of random numbers.

44. a. Normal distribution with

$E(\bar{x}) = \mu = 406$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{64}} = 10$

Since $n/N = 64/3400 = .0188 < .05$; therefore, the finite population correction factor is not necessary.

b. $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{15}{80 / \sqrt{64}} = 1.50 \quad P(z \leq 1.50) = .9332$

$P(z < -1.50) = .0668$

$P(391 \leq \bar{x} \leq 421) = P(-1.50 \leq z \leq 1.50) = .9332 - .0668 = .8664$

Using Excel: $\text{NORM.DIST}(421,406,80/\text{SQRT}(64),\text{TRUE}) - \text{NORM.DIST}(391,406,80/\text{SQRT}(64),\text{TRUE}) = .8664$

c. At $\bar{x} = 380, \quad z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{380 - 406}{80 / \sqrt{64}} = -2.60$

$P(\bar{x} \leq 380) = P(z \leq -2.60) = .0047$

Using Excel: $\text{NORM.DIST}(380,406,80/\text{SQRT}(64),\text{TRUE}) = .0047$
Yes, this is an unusually low performing group of 64 stores. The probability of a sample mean annual sales per square foot of $380 or less is only .0047.

46. \( \mu = 27,175 \quad \sigma = 7400 \)

a. \( \sigma_x = \frac{7400}{\sqrt{64}} = 955 \)

b. \( z = \frac{\bar{x} - \mu}{\sigma_x} = \frac{0}{955} = 0 \)

\( P(\bar{x} > 27,175) = P(z > 0) = .50 \)

Note: This could have been answered easily without any calculations; 27,175 is the expected value of the sampling distribution of \( \bar{x} \).
Chapter 7

c. \[ z = \frac{\bar{x} - \mu}{\sigma_x} = \frac{1000}{955} = 1.05 \] 
\[ P(\bar{x} \leq 1.05) = .8531 \]

\[ P(\bar{x} < -1.05) = .1469 \]

Using Excel: \[ \text{NORM.DIST}(28175,27175,7400/\sqrt{60},\text{TRUE}) - \text{NORM.DIST}(26175,27175,7400/\sqrt{60},\text{TRUE}) = .7048 \]

d. \[ \sigma_x = \frac{7400}{\sqrt{100}} = 740 \]
\[ z = \frac{\bar{x} - \mu}{\sigma_x} = \frac{1000}{740} = 1.35 \] 
\[ P(\bar{x} \leq 1.35) = .9115 \]

\[ P(\bar{x} < -1.35) = .0885 \]

Using Excel: \[ \text{NORM.DIST}(28175,27175,7400/\sqrt{100},\text{TRUE}) - \text{NORM.DIST}(26175,27175,7400/\sqrt{100},\text{TRUE}) = .8234 \]

48. a. \[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{20}} = 20 \]
\[ \sqrt{n} = \frac{500}{20} = 25 \] and \( n = (25)^2 = 625 \)

b. For ± 25,
\[ z = \frac{25}{20} = 1.25 \] 
\[ P(\bar{x} \leq 1.25) = .8944 \]

\[ P(\bar{x} < -1.25) = .1056 \]

Probability = \[ P(-1.25 \leq \bar{x} \leq 1.25) = .8944 - .1056 = .7888 \]

Using Excel: \[ \text{NORM.S.DIST}(25/20,\text{TRUE}) - \text{NORM.S.DIST}(-25/20,\text{TRUE}) = .7887 \]

50. \( p = .15 \)

a. This is the graph of a normal distribution with \( E(\bar{p}) = p = .15 \) and
\[ \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.15(1-.15)}{240}} = .0230 \]

The normal distribution is appropriate because \( np = 240(.15) = 36 \) and \( n(1 - p) = 240(.85) = 204 \) are both greater than 5.
b. Within ± .04 means .11 ≤ \( \bar{p} \) ≤ .19

\[
\begin{align*}
z &= \frac{\bar{p} - p}{\sigma_p} = \frac{.19 - .15}{.0230} = 1.74 \\
&= \frac{.11 - .15}{.0230} = -1.74
\end{align*}
\]

\( P(.11 \leq \bar{p} \leq .19) = P(-1.74 \leq z \leq 1.74) = .9591 - .0409 = .9182 \)

Using Excel: \( \text{NORM.DIST}(.19,.15,\text{SQRT}(.15*.85/240),\text{TRUE}) - \text{NORM.DIST}(.11,.15,\text{SQRT}(.15*.85/240),\text{TRUE}) = .9173 \)

c. Within ± .02 means .13 ≤ \( \bar{p} \) ≤ .17

\[
\begin{align*}
z &= \frac{\bar{p} - p}{\sigma_p} = \frac{.17 - .15}{.0230} = .87 \\
&= \frac{.13 - .15}{.0230} = -.87
\end{align*}
\]

\( P(.13 \leq \bar{p} \leq .17) = P(-.87 \leq z \leq .87) = .8078 - .1922 = .6156 \)

Using Excel: \( \text{NORM.DIST}(.17,.15,\text{SQRT}(.15*.85/240),\text{TRUE}) - \text{NORM.DIST}(.13,.15,\text{SQRT}(.15*.85/240),\text{TRUE}) = .6145 \)

52. a.

\[
\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.40)(1-.40)}{380}} = .0251
\]

Within ± .04 means .36 ≤ \( \bar{p} \) ≤ .44

\[
\begin{align*}
z &= \frac{.44 - .40}{.0251} = 1.59 \\
z &= \frac{.36 - .40}{.0251} = -1.59
\end{align*}
\]

\( P(.36 \leq \bar{p} \leq .44) = P(-1.59 \leq z \leq 1.59) = .9441 - .0559 = .8882 \)

Using Excel: \( \text{NORM.DIST}(.44,.40,\text{SQRT}(.4*.6/380),\text{TRUE}) - \text{NORM.DIST}(.36,.40,\text{SQRT}(.4*.6/380),\text{TRUE}) = .8885 \)

b. We want \( P(\bar{p} \geq .45) \)

\[
\begin{align*}
z &= \frac{\bar{p} - p}{\sigma_p} = \frac{.45 - .40}{.0251} = 1.99
\end{align*}
\]

\( P(\bar{p} \geq .45) = P(z \geq 1.99) = 1 - P(\bar{p} < .45) = 1 - P(z < 1.99) = 1 - .9767 = .0233 \)

Using Excel: \( 1 - \text{NORM.DIST}(.45,.4,\text{SQRT}(.4*.6/380),\text{TRUE}) = .0233 \)

54. a.

\[
\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{25(.75)}{n}} = .0625
\]

Solve for \( n \)
\[ n = \frac{.25(.75)}{(0.0625)^2} = 48 \]

b. Normal distribution with \( E(\bar{p}) = p = .25 \) and \( \sigma_{\bar{p}} = .0625 \)

(Note: (48)(.25) = 12 > 5, and (48)(.75) = 36 > 5)

c. \( P(\bar{p} \geq .30) = ? \)

\[ z = \frac{.30 - .25}{.0625} = .80 \]

\[ P(z \leq .80) = .7881 \]

\[ P(\bar{p} \geq .30) = 1 - .7881 = .2119 \]

Using Excel: 1-NORM.DIST(.30,.25,.0625,TRUE) = .2119
Chapter 8
Interval Estimation

Solutions:

2. a. $32 \pm 1.645 \left( \frac{6}{\sqrt{50}} \right)$
   
   $32 \pm 1.4$ or 30.6 to 33.4
   Using Excel for the margin of error: CONFIDENCE.NORM(.10,6,50) = 1.40

   b. $32 \pm 1.96 \left( \frac{6}{\sqrt{50}} \right)$
   
   $32 \pm 1.66$ or 30.34 to 33.66
   Using Excel for the margin of error: CONFIDENCE.NORM(.05,6,50) = 1.66

   c. $32 \pm 2.576 \left( \frac{6}{\sqrt{50}} \right)$
   
   $32 \pm 2.19$ or 29.81 to 34.19
   Using Excel for the margin of error: CONFIDENCE.NORM(.01,6,50) = 2.19

4. Sample mean $\bar{x} = \frac{160 + 152 + 156}{2} = 156$
   
   Margin of Error = $160 - 156 = 4$
   
   $1.96\left( \frac{\sigma}{\sqrt{n}} \right) = 4$
   
   $\sqrt{n} = \frac{1.96\sigma}{4} = \frac{1.96(15)}{4} = 7.35$
   
   $n = (7.35)^2 = 54$

5. Using Excel and the webfile Houston, the sample mean is $\bar{x} = 21.52$ and the sample size is $n = 64$.
   The population standard deviation $\sigma = 6$ is known.
   
   a. With 99% confidence $z_{0.025} = z_{0.05} = 2.576$
   
   Margin of Error = $2.576\sigma / \sqrt{n} = 2.576(6 / \sqrt{64}) = 1.93$
   Using Excel for the margin of error: CONFIDENCE.NORM(.01,6,64) = 1.93
   
   b. Confidence Interval: $21.52 \pm 1.93$ or 19.59 to 23.45

6. A 95% confidence interval is of the form

   $\bar{x} \pm z_{0.025} \left( \frac{\sigma}{\sqrt{n}} \right)$

   Using Excel and the webfile TravelTax, the sample mean is $\bar{x} = 40.31$ and the sample size is $n = 200$.
   The population standard deviation $\sigma = 8.5$ is known. The confidence interval is
40.31 ± 1.96(8.5/\sqrt{200})

40.31 ± 1.18 or 39.13 to 41.49
Using Excel for the margin of error: CONFIDENCE.NORM(.05,8.5,200) = 1.18

8. a. Since \( n \) is small, an assumption that the population is at least approximately normal is required so that the sampling distribution of \( \bar{x} \) can be approximated by a normal distribution.

b. Margin of error: \( z_{0.025} \left( \frac{\sigma}{\sqrt{n}} \right) = 1.96(\sqrt{5.5/10}) = 3.41 \)
Using Excel for the margin of error: CONFIDENCE.NORM(.05,5.5,10) = 3.41

c. Margin of error: \( z_{0.05} \left( \frac{\sigma}{\sqrt{n}} \right) = 2.576(\sqrt{5.5/10}) = 4.48 \)
Using Excel for the margin of error: CONFIDENCE.NORM(.01,5.5,10) = 4.48

10. a. \( \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \)

3486 ± 1.645 (650/\sqrt{120})

3486 ± 98 or $3388 to $3584
Using Excel for the margin of error: CONFIDENCE.NORM(.10,650,120) = 98

b. 3486 ± 1.96 (650/\sqrt{120})

3486 ± 116 or $3370 to $3602
Using Excel for the margin of error: CONFIDENCE.NORM(.05,650,120) = 116

c. 3486 ± 2.576 (650/\sqrt{120})

3486 ± 153 or $3333 to $3639
Using Excel for the margin of error: CONFIDENCE.NORM(.01,650,120) = 153

d. The confidence interval gets wider as we increase our confidence level. We need a wider interval to be more confident that the interval will contain the population mean.

12. a. 2.179

b. -1.676

c. 2.457

d. Use .05 column, -1.708 and 1.708

e. Use .025 column, -2.014 and 2.014

13. \( n=8 \text{df} = 7 \)

a. \( \bar{x} = \frac{\sum x_i}{n} = \frac{80}{8} = 10 \)
b. 

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$(x_i - \bar{x})$</th>
<th>$(x_i - \bar{x})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>-4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>-5</td>
<td>25</td>
</tr>
</tbody>
</table>

\[ s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{84}{7}} = 3.464 \]

c. \[ t_{0.025}(s/\sqrt{n}) = 2.365(3.464/\sqrt{8}) = 2.9 \]

d. \[ \bar{x} \pm t_{0.025}(s/\sqrt{n}) \]
\[ 10 \pm 2.9 \text{ or 7.1 to 12.9} \]

14. \[ n=54 \quad s = 4.4 \quad \bar{x} \pm t_{n/2}(s/\sqrt{n}) \] \[ df = 53 \]

a. \[ 22.5 \pm 1.674(4.4/\sqrt{54}) \]
\[ 22.5 \pm 1 \text{ or 21.5 to 23.5} \]

b. \[ 22.5 \pm 2.006(4.4/\sqrt{54}) \]
\[ 22.5 \pm 1.2 \text{ or 21.3 to 23.7} \]

c. \[ 22.5 \pm 2.672(4.4/\sqrt{54}) \]
\[ 22.5 \pm 1.6 \text{ or 20.9 to 24.1} \]

d. As the confidence level increases, there is a larger margin of error and a wider confidence interval.

15. \[ n=65 \quad s = 5.2 \quad \bar{x} \pm t_{n/2}(s/\sqrt{n}) \]

90% confidence \[ df = 64 \quad t_{0.05} = 1.669 \]
\[ 19.5 \pm 1.669(5.2/\sqrt{65}) \]
\[ 19.5 \pm 1.08 \text{ or 18.42 to 20.58} \]

95% confidence \[ df = 64 \quad t_{0.025} = 1.998 \]
\[ 19.5 \pm 1.998(5.2/\sqrt{65}) \]
\[ 19.5 \pm 1.29 \text{ or 18.21 to 20.79} \]
16. a. For the CorporateBonds data set, the output obtained using Excel’s Descriptive Statistics tool for Years to Maturity follows:

<table>
<thead>
<tr>
<th>Years to Maturity</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>
<tr>
<td>Confidence Level(95.0%)</td>
</tr>
</tbody>
</table>

Using Excel, $\bar{x} = 9.71$ and $s = 7.98$

The sample mean years to maturity is 9.71 years with a standard deviation of 7.98.

b. $\bar{x} \pm t_{0.025} \left( \frac{s}{\sqrt{n}} \right)$ $n=40$        $df = 39$        $t_{0.025} = 2.023$

$9.71 \pm 2.023 \left( \frac{7.98}{\sqrt{40}} \right)$

$9.71 \pm 2.55$ or 7.15 to 12.26

The 95% confidence interval for the population mean years to maturity is 7.15 to 12.26 years.

c. For the CorporateBonds data set, the output obtained using Excel’s Descriptive Statistics tool for Yield follows:

<table>
<thead>
<tr>
<th>Yield</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>
<tr>
<td>Confidence Level(95.0%)</td>
</tr>
</tbody>
</table>

Using Excel, $\bar{x} = 3.8854$ and $s = 1.6194$
The sample mean yield on corporate bonds is 3.8854% with a standard deviation of 1.6194.

d. \( \bar{x} \pm t_{0.025} \left( \frac{s}{\sqrt{n}} \right) \)

\( df = 39 \quad t_{0.025} = 2.023 \)

\[ 3.8854 \pm 2.023 \left( \frac{1.6194}{\sqrt{40}} \right) \]

\[ 3.8854 \pm .5180 \text{ or } 3.3674 \text{ to } 4.4033 \]

The 95% confidence interval for the population mean yield is 3.3674 to 4.4033 percent.

18. For the JobSearch data set, the output obtained using Excel’s Descriptive Statistics tool follows:

<table>
<thead>
<tr>
<th>Job Search Time (Weeks)</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>
<tr>
<td>Confidence Level(95.0%)</td>
</tr>
</tbody>
</table>

a. \( \bar{x} = 22 \text{ weeks} \)

b. margin of error = 3.8014

c. The 95% confidence interval is \( \bar{x} \pm \text{margin of error} \)

\[ \bar{x} \pm t_{0.025} \left( s / \sqrt{n} \right) \text{ with } n=40 \quad df = 39 \quad t=2.023 \]

\[ 22 \pm 2.023 \left( \frac{11.8862}{\sqrt{40}} \right) \]

\[ 22 \pm 3.8014 \text{ or } 18.20 \text{ to } 25.80 \]

d. The descriptive statistics output shows that the skewness is 1.0062. There is a moderate positive skewness in this data set. This can be expected to exist in a population such as this. While the above results are acceptable, considering a slightly larger sample next time would be a good strategy.
20. a. For the AutoInsurance data set, the output obtained using Excel’s Descriptive Statistics tool follows:

<table>
<thead>
<tr>
<th>Annual Premium</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>

The point estimate of the mean annual auto insurance premium in Michigan is $2551.

b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum (x_i - 2551)^2}{20-1}} = 301.3077$

95% confidence interval: $\bar{x} \pm t_{0.025} \left( \frac{s}{\sqrt{n}} \right) \quad df = 19$

$2551 \pm 2.093 \left( \frac{301.3077}{\sqrt{20}} \right)$

$2551 \pm 141.01$ or $2409.99$ to $2692.01$

c. The 95% confidence interval for Michigan does not include the national average of $1503 for the United States. We would be 95% confident that auto insurance premiums in Michigan are above the national average.

22. a. The Excel output from using the Descriptive Statistics analysis tool with the Guardians file is shown:

<table>
<thead>
<tr>
<th>Revenue (S)</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>
Interval Estimation

Confidence Level (95.0%)  1486.86387

\[ x \pm t_{0.025} \left( \frac{s}{\sqrt{n}} \right) \quad \text{with } n=30 \quad df = 29 \quad s=3981.89 \quad t=2.045 \]

\[
23100 \pm 2.045 \left( \frac{3981.89}{\sqrt{30}} \right)
\]

\[
23100 \pm 14686.86
\]

The sample mean is 23,100 and the margin of error (Confidence Level) is 1486.86.

The 95% confidence interval for the population mean is $21,613.14 to $24,586.86. We are 95% confident that the population mean two-day ticket sales revenue per theater is between $21,613.14 and $24,586.86.

b. Mean number of customers per theater = \[
\frac{23,100}{8.11} = 2848.34
\]

c. Total number of customers = \[
4080(2848.34) = 11,621,227 \text{ (or } 11,621,208 \text{ if calculated from the unrounded prior calculations)} \approx $11.6 million customers
\]

Total box office ticket sales for the three-day weekend = \[
4080(23,100) = $94,248,000 \approx $94 million
\]

24. a. Planning value of \( \sigma = \frac{\text{Range}}{4} = \frac{36}{4} = 9 \)

b. \[
n = \frac{z_{0.025}^2 \sigma^2}{E^2} = \frac{(1.96)^2 (9)^2}{3^2} = 34.57 \quad \text{Use } n = 35
\]

c. \[
n = \frac{(1.96)^2 (9)^2}{2^2} = 77.79 \quad \text{Use } n = 78
\]

25. a. \[
n = \frac{(1.96)^2 (6.84)^2}{(1.5)^2} = 79.88 \quad \text{Use } n = 80
\]

b. \[
n = \frac{(1.645)^2 (6.84)^2}{2^2} = 31.65 \quad \text{Use } n = 32
\]

26. a. \[
n = \frac{z_{0.025}^2 \sigma^2}{E^2} = \frac{(1.96)^2 (.25)^2}{(.10)^2} = 24.01 \quad \text{Use } 25
\]

If the normality assumption for the population appears questionable, this should be adjusted upward to at least 30.

b. \[
n = \frac{(1.96)^2 (.25)^2}{(.07)^2} = 49 \quad \text{Use 49 to guarantee a margin of error no greater than .07. However, the US EIA may choose to increase the sample size to a round number of 50}
\]

c. \[
n = \frac{(1.96)^2 (.25)^2}{(.05)^2} = 96.04 \quad \text{Use 97}
\]

For reporting purposes, the US EIA might decide to round up to a sample size of 100.
28. a. \[ n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{(1.645)^2(25)^2}{(3)^2} = 187.92 \] Use \( n = 188 \)

b. \[ n = \frac{(1.96)^2(25)^2}{(3)^2} = 266.78 \] Use \( n = 267 \)

c. \[ n = \frac{(2.576)^2(25)^2}{(3)^2} = 460.82 \] Use \( n = 461 \)

d. The sample size gets larger as the confidence level is increased. We would not recommend 99% confidence. The sample size must be increased by 79 respondents \((267 - 188)\) to go from 90% to 95%. This may be reasonable. However, increasing the sample size by 194 respondents \((461 - 267)\) to go from 95% to 99% would probably be viewed as too expensive and time consuming for the 4% gain in confidence level.

30. \( \sigma = 1.5 \)

\[ n = \frac{z_{0.025}^2 \sigma^2}{E^2} = \frac{(1.96)^2(1.5)^2}{(.5)^2} = 34.57 \]

Use \( n = 35 \) to guarantee the margin of error will not exceed .5.

31. a. \( p = \frac{100}{400} = .25 \)

b. \[ \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.25(.75)}{400}} = .0217 \]

c. \[ p \pm z_{0.025} \sqrt{\frac{p(1-p)}{n}} \]
\[ .25 \pm 1.96 (.0217) \]
\[ .25 \pm .0424 \] or .2076 to .2924

32. a. \[ .70 \pm 1.645 \sqrt{\frac{.70(.30)}{800}} \]
\[ .70 \pm .0267 \] or .6733 to .7267

b. \[ .70 \pm 1.96 \sqrt{\frac{.70(.30)}{800}} \]
\[ .70 \pm .0318 \] or .6682 to .7318

34. Use planning value \( p = .50 \)

\[ n = \frac{(1.96)^2(.50)(.50)}{(0.03)^2} = 1067.11 \] Use \( n = 1068 \)
35. a. \( \hat{p} = \frac{20901}{45535} = 0.459 \)

b. Margin of Error:

\[
Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645 \sqrt{\frac{0.459(0.541)}{45535}} = 0.0038
\]

c. Confidence interval:

\( 0.459 \pm 0.0038 \) or \( 0.4552 \) to \( 0.4628 \) (without any intermediary rounding)

d. Margin of Error

\[
Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \sqrt{\frac{0.459(0.541)}{45535}} = 0.0046
\]

95% Confidence Interval

\( 0.459 \pm 0.0046 \) or \( 0.4544 \) to \( 0.4636 \)

36. a. \( \hat{p} = \frac{46}{200} = 0.23 \)

b. \( \hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

\( 0.23 \pm 1.96 \sqrt{\frac{(0.23)(0.77)}{200}} \)

\( 0.23 \pm 0.0583 \) or \( 0.1717 \) to \( 0.2883 \)

38. a. \( \hat{p} = \frac{29}{162} = 0.1790 \)

b. \( \hat{p} = \frac{104}{162} = 0.6420 \)

Margin of error = \( 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \sqrt{\frac{(0.642)(0.358)}{162}} = 0.0738 \)

Confidence interval: \( 0.6420 \pm 0.0738 \) or \( 0.5682 \) to \( 0.7158 \)

c. \( n = \frac{1.96^2(0.642)(0.358)}{(0.05)^2} = 353.18 \) Use \( n = 354 \)

39. a. \( n = \frac{z^2_{0.025}p^*(1-p^*)}{E^2} = \frac{(1.96)^2(0.17)(1-0.17)}{(0.03)^2} = 602.28 \) Use \( n = 603 \)

b. \( n = \frac{z^2_{0.01}p^*(1-p^*)}{E^2} = \frac{(2.576)^2(0.17)(1-0.17)}{(0.03)^2} = 1040.34 \) Use \( n = 1041 \)

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
40. Margin of error: 
\[ z_{0.05} \sqrt{\frac{p(1-p)}{n}} = 1.96 \sqrt{\frac{.52(1-.52)}{800}} = .0346 \]

95% Confidence interval: 
\[ \bar{p} \pm .0346 \]

.52 ± .0346 or .4854 to .5546

42. a. \[ \sqrt{\frac{p\cdot(1-p)}{n}} = \sqrt{\frac{.50(1-.50)}{491}} = .0226 \]

\[ z_{0.05} \sqrt{\frac{p\cdot(1-p)}{n}} = 1.96(.0226) = .0442 \]

b. \[ n = \frac{z_{0.05}^2 \cdot p\cdot(1-p)}{E^2} \]

September \[ n = \frac{1.96^2 \cdot (.50)(1-.50)}{.04^2} = 600.25 \] Use 601

October \[ n = \frac{1.96^2 \cdot (.50)(1-.50)}{.03^2} = 1067.11 \] Use 1068

November \[ n = \frac{1.96^2 \cdot (.50)(1-.50)}{.02^2} = 2401 \]

Pre-Election \[ n = \frac{1.96^2 \cdot (.50)(1-.50)}{.01^2} = 9604 \]

44. Using Excel and the webfile FedTaxErrors, the sample mean is \( \bar{x} = 326.6674 \) and the sample size is \( n = 10001 \). The population standard deviation \( \sigma = 12300 \) is known.

a. \( \bar{x} = 326.6674 \)

b. \( z_{a/2}\bar{x} = 1.96(.122.9939) = 241.0680 \)

Using Excel for the margin of error: \( \text{CONFIDENCE.NORM(.05,12300,10001)} = 241.0635 \)

c. \( \bar{x} \pm z_{a/2}\sigma_{\bar{x}} = 326.6674 \pm 241.0680 = (85.5994, 567.7354) \) or 85.6 to 567.7

46. a. \( \bar{p} = 65120/102519 = .6352 \)

\[ \sigma_p = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{.6352(1-.6352)}{102519}} = .0015 \]

b. \( z_{a/2}\bar{p} = 1.96(.0015) = .0029 \)

c. \( \bar{p} \pm z_{a/2}\sigma_p = .6352 \pm .0029 = (.6323, .6381) \)

48. a. Margin of error: 
\[ z_{0.05} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{15}{\sqrt{54}} = 4.00 \]

Using Excel for the margin of error: \( \text{CONFIDENCE.NORM(.05,15,54)} = 4.0 \)

b. Confidence interval: \( \bar{x} \pm \text{margin of error} \)
Interval Estimation

$$33.77 \pm 4.00 \text{or } \$29.77 \text{ to } \$37.77$$

50. \( \bar{x} = 1873 \quad n = 80 \)

a. Margin of error = \( t_{0.025} \left( \frac{s}{\sqrt{n}} \right) \)

\( df = 79t_{0.025} = 1.990s = 550 \)

\( 1.990 \left( \frac{550}{\sqrt{80}} \right) = 122 \)

b. \( \bar{x} \pm \text{margin of error} \)

\( 1873 \pm 122 \text{ or } \$1751 \text{ to } \$1995 \)

c. 92 million Americans are of age 50 and over

Estimate of total expenditures = \( 92(1873) = 172,316 \)

In dollars, we estimate that 172,316 million dollars are spent annually by Americans of age 50 and over on restaurants and carryout food.

d. We would expect the median to be less than the mean. The few individuals that spend much more than the average cause the mean to be larger than the median. This is typical for data of this type.

52. a. For the DrugCost data set, the output obtained using Excel’s Descriptive Statistics tool for Total Annual Cost follows:

<table>
<thead>
<tr>
<th>Total Annual Cost</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>773</td>
</tr>
<tr>
<td>Standard Error</td>
<td>36.91917</td>
</tr>
<tr>
<td>Median</td>
<td>647</td>
</tr>
<tr>
<td>Mode</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>738.3835</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>545210.1</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.65717</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.577692</td>
</tr>
<tr>
<td>Range</td>
<td>3366</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>3366</td>
</tr>
<tr>
<td>Sum</td>
<td>309200</td>
</tr>
<tr>
<td>Count</td>
<td>400</td>
</tr>
<tr>
<td>Confidence Level(90.0%)</td>
<td>60.86796</td>
</tr>
</tbody>
</table>

\( \bar{x} = 773 \text{ and } s = 738.3835 \quad n = 400 \quad df = 399 \)

\( t = 1.6487 \quad \text{This can be obtained from Excel using: } T.INV(.05,399) = 1.6487 \)

Margin of error = \( t_{0.05} \left( \frac{s}{\sqrt{n}} \right) = 1.6487 \left( \frac{738.3835}{\sqrt{400}} \right) = 60.87 \)
90% Confidence Interval: 773 ± 60.87 or $712.13 to $833.87

b. For the DrugCost data set, the output obtained using Excel’s Descriptive Statistics tool for Employee Out-of-Pocket Cost follows:

<table>
<thead>
<tr>
<th>Employee Out-of-Pocket Cost</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>
<tr>
<td>Confidence Level (90.0%)</td>
</tr>
</tbody>
</table>

\[
\bar{x} = 187 \quad \text{and} \quad s = 178.6207 \quad n=400 \quad df=399 \quad t= 1.6487 \quad \text{obtained from T.INV(.05,399)}
\]

Margin of error = \( t_{.05} \left( \frac{s}{\sqrt{n}} \right) = 1.6487 \left( \frac{178.6207}{\sqrt{400}} \right) = 14.72 \)

90% Confidence Interval: 187 ± 14.72 or $172.28 to $201.72

c. There were 136 employees who had no prescription medication cost for the year.
   \[ \bar{p} = \frac{136}{400} = .34 \]

d. The margin of error in part (a) is 60.87; the margin of error in part (c) is 14.72. The margin of error in part (a) is larger because the sample standard deviation in part (a) is larger. The sample size and confidence level are the same in both parts.
60. a. With 165 out of 750 respondents rating the economy as good or excellent,
\[ \bar{p} = \frac{165}{750} = 0.22 \]

b. Margin of error
\[
= 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96 \sqrt{\frac{0.22(1-0.22)}{750}} = 0.0296
\]
95% Confidence interval: 0.22 ± 0.0296 or 0.1904 to 0.2496

c. With 315 out of 750 respondents rating the economy as poor,
\[ \bar{p} = \frac{315}{750} = 0.42 \]

Margin of error
\[
= 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96 \sqrt{\frac{0.42(1-0.42)}{750}} = 0.0353
\]
95% Confidence interval: 0.42 ± 0.0353 or 0.3847 to 0.4553

d. The confidence interval in part (c) is wider. This is because the sample proportion is closer to 0.5 in part (c).

62. a. 
\[
n = \frac{(2.33)^2(0.70)(0.30)}{(0.03)^2} = 1267.74 \quad \text{Use } n = 1267
\]

b. 
\[
n = \frac{(2.33)^2(0.50)(0.50)}{(0.03)^2} = 1508.03 \quad \text{Use } n = 1509
\]

64. a. 
\[ n = 1993 \quad \bar{p} = \frac{618}{1993} = 0.3101 \]

b. 
\[ \bar{p} \pm 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.3101 \pm 1.96 \sqrt{\frac{(0.3101)(0.6899)}{1993}} = 0.3101 \pm 0.0203 \text{ or } 0.2898 \text{ to } 0.3304
\]

c. 
\[
n = \frac{z^2 \bar{p}*(1-\bar{p}*)}{E^2}
\]
\[ z = \frac{(1.96)^2(0.3101)(0.6899)}{(0.01)^2} = 8218.64 \quad \text{Use } n = 8219
\]

No; the sample appears unnecessarily large. The .02 margin of error reported in part (b) should provide adequate precision.
Chapter 9
Hypothesis Tests

Solutions:

2.  a.  \( H_0: \mu \leq 14 \)
\( H_\alpha: \mu > 14 \)  Research hypothesis

b.  There is no statistical evidence that the new bonus plan increases sales volume.

c.  The research hypothesis that \( \mu > 14 \) is supported. We can conclude that the new bonus plan increases the mean sales volume.

4.  a.  \( H_0: \mu \geq 220 \)
\( H_\alpha: \mu < 220 \)  Research hypothesis to see if mean cost is less than $220.

b.  We are unable to conclude that the new method reduces costs.

c.  Conclude \( \mu < 220 \). Consider implementing the new method based on the conclusion that it lowers the mean cost per hour.

5.  a.  Conclude that the population mean monthly cost of electricity in the Chicago neighborhood is greater than $104 and hence higher than in the comparable neighborhood in Cincinnati.

b.  The Type I error is rejecting \( H_0 \) when it is true. This error occurs if the researcher concludes that the population mean monthly cost of electricity is greater than $104 in the Chicago neighborhood when the population mean cost is actually less than or equal to $104.

c.  The Type II error is accepting \( H_0 \) when it is false. This error occurs if the researcher concludes that the population mean monthly cost for the Chicago neighborhood is less than or equal to $104 when it is not.

6.  a.  \( H_0: \mu \leq 1 \)  The label claim or assumption.
\( H_\alpha: \mu > 1 \)

b.  Claiming \( \mu > 1 \) when it is not. This is the error of rejecting the product’s claim when the claim is true.

c.  Concluding \( \mu \leq 1 \) when it is not. In this case, we miss the fact that the product is not meeting its label specification.
8. a. \( H_0: \mu \geq 220 \quad \) Research hypothesis to see if new method reduces the operating cost/hr.

\( H_1: \mu < 220 \)

b. Claiming \( \mu < 220 \) when the new method does not lower costs. A mistake could be implementing the method when it does not help.

c. Concluding \( \mu \geq 220 \) when the method really would lower costs. This could lead to not implementing a method that would lower costs.

10. a. 
\[
\begin{align*}
  z &= \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{26.4 - 25}{6 / \sqrt{40}} = 1.48
\end{align*}
\]

b. \( p \)-value is the area in the upper tail: \( P(z \geq 1.48) \)

Using normal table with \( z = 1.48 \): \( p \)-value = 1.0000 - .9306 = .0694

Using Excel: \( p \)-value = 1 - NORM.S.DIST(1.48,TRUE) = .0694

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0700

c. \( p \)-value > .01, do not reject \( H_0 \)

d. Reject \( H_0 \) if \( z \geq 2.33 \)

1.48 < 2.33, do not reject \( H_0 \)

12. a. 
\[
\begin{align*}
  z &= \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{78.5 - 80}{12 / \sqrt{100}} = -1.25
\end{align*}
\]

\( p \)-value is the lower-tail area

Using normal table with \( z = -1.25 \): \( p \)-value = .1056

Using Excel: \( p \)-value = NORM.S.DIST(-1.25,TRUE) = .1056

\( p \)-value > .01, do not reject \( H_0 \)

b. 
\[
\begin{align*}
  z &= \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{77 - 80}{12 / \sqrt{100}} = -2.50
\end{align*}
\]

\( p \)-value is the lower-tail area

Using normal table with \( z = -2.50 \): \( p \)-value = .0062

Using Excel: \( p \)-value = NORM.S.DIST(-2.50,TRUE) = .0062

\( p \)-value \leq .01, reject \( H_0 \)

c. 
\[
\begin{align*}
  z &= \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{75.5 - 80}{12 / \sqrt{100}} = -3.75
\end{align*}
\]

\( p \)-value is the lower-tail area

Using normal table with \( z = -3.75 \): \( p \)-value \approx 0

Using Excel: \( p \)-value = NORM.S.DIST(-3.75,TRUE) \approx .0001 \approx 0

\( p \)-value \leq .01, reject \( H_0 \)
d. \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{81 - 80}{12/\sqrt{100}} = .83 \]

\( p \)-value is the lower-tail area

Using normal table with \( z = .83 \): \( p \)-value = .7967
Using Excel: \( p \)-value = NORM.S.DIST(.83,TRUE) = .7967
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .7977

\( p \)-value > .01, do not reject \( H_0 \)

14. a. \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{23 - 22}{10/\sqrt{75}} = .87 \]

Because \( z > 0 \), \( p \)-value is two times the upper tail area

Using normal table with \( z = .87 \): \( p \)-value = 2(1 - .8078) = .3844
Using Excel: \( p \)-value = 2*(1-NORM.S.DIST(.87,TRUE)) = .3843
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .3865

\( p \)-value > .01, do not reject \( H_0 \)

b. \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{25.1 - 22}{10/\sqrt{75}} = 2.68 \]

Because \( z > 0 \), \( p \)-value is two times the upper tail area

Using normal table with \( z = 2.68 \): \( p \)-value = 2(1 - .9963) = .0074
Using Excel: \( p \)-value = 2*(1-NORM.S.DIST(2.68,TRUE)) = .0074
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0073

\( p \)-value \( \leq \) .01, reject \( H_0 \)

c. \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{20 - 22}{10/\sqrt{75}} = -1.73 \]

Because \( z < 0 \), \( p \)-value is two times the lower tail area

Using normal table with \( z = -1.73 \): \( p \)-value = 2(.0418) = .0836
Using Excel: \( p \)-value = 2*(NORM.S.DIST(-1.73,TRUE)) = .0836
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0833

\( p \)-value > .01, do not reject \( H_0 \)

15. a. \( H_0: \mu \geq 1056 \)
\( H_a: \mu < 1056 \)    Research hypothesis

b. \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{910 - 1056}{1600/\sqrt{400}} = -1.83 \]

\( p \)-value is the lower-tail area

Using normal table with \( z = -1.83 \): \( p \)-value = .0336
Using Excel: \( p \)-value = NORM.S.DIST(-1.83,TRUE) = .0336
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0340
c. $p$-value $\leq .05$, reject $H_0$. Conclude the mean refund of “last minute” filers is less than $1056.$

d. Reject $H_0$ if $z \leq -1.645$

-1.83 $\leq -1.645$, reject $H_0$

16. a. $H_0$: $\mu \leq 3173$

$H_a$: $\mu > 3173$  

Research hypothesis

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{3325 - 3173}{1000 / \sqrt{180}} = 2.04$

$p$-value is the upper tail area or $P(z \geq 2.04)$

Using normal table with $z = 2.04$; $p$-value = $1.0000 - .9793 = .0207$

Using Excel: $p$-value = $1$-$\text{NORM.S.DIST}(2.04, TRUE) = .0207$

Using unrounded Test Statistic via Excel with cell referencing, $p$-value = .0207

c. $p$-value $\leq .05$. Reject $H_0$. The current population mean credit card balance for undergraduate

students has increased compared to the previous all-time high of $3173$ reported in April 2009.

18. a. $H_0$: $\mu = 192$

$H_a$: $\mu \neq 192$  

Research hypothesis

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{182 - 192}{55 / \sqrt{150}} = -2.23$

Because $z < 0$, $p$-value is two times the lower tail area

Using normal table with $z = -2.23$; $p$-value = $2 \times .0129 = .0258$

Using Excel: $p$-value = $2 \times \text{NORM.S.DIST}(-2.23, TRUE) = .0258$

Using unrounded Test Statistic via Excel with cell referencing, $p$-value = .0260

c. $p$-value = $0.0258 \leq \alpha = .05$

Reject $H_0$ and conclude that the mean number of restaurant meals eaten by young millennials has

changed in 2012.

20. a. $H_0$: $\mu \geq 838$

$H_a$: $\mu < 838$  

Research hypothesis

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{745 - 838}{300 / \sqrt{60}} = -2.40$

c. Lower tail $p$-value is area to left of the test statistic.

Using normal table with $z = -2.40$; $p$-value = .0082.

Using Excel: $p$-value = $\text{NORM.S.DIST}(-2.40, TRUE) = .0082$

Using unrounded Test Statistic via Excel with cell referencing, $p$-value = .0082

d. $p$-value $\leq .01$; reject $H_0$. Conclude that the annual expenditure per person on prescription drugs is

less in the Midwest than in the Northeast.
22. a. \( H_0: \mu = 8 \)
\( H_a: \mu \neq 8 \)  
Research hypothesis

b. \[
\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{8.4 - 8.0}{3.2 / \sqrt{120}} = 1.37
\]

Because \( z > 0 \), \( p \)-value is two times the upper tail area

Using normal table with \( z = 1.37 \): \( p \)-value = \( 2(1 - .9147) = .1706 \)

Using Excel: \( p \)-value = \( 2*(1-NORM.S.DIST(1.37,TRUE)) = .1707 \)

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = \( .1709 \)

c. \( p \)-value > .05, do not reject \( H_0 \). Cannot conclude that the population mean waiting time differs from 8 minutes.

d. \( \bar{x} \pm z_{.025}(\sigma / \sqrt{n}) \)
\[ 8.4 \pm 1.96 \left(3.2 / \sqrt{120}\right) \]
\[ 8.4 \pm .57 \quad (7.83 \text{ to } 8.97) \]

Yes; \( \mu = 8 \) is in the interval. Do not reject \( H_0 \).

24. a. \[
t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{17 - 18}{4.5 / \sqrt{48}} = -1.54
\]

b. Degrees of freedom = \( n - 1 = 47 \)

Because \( t < 0 \), \( p \)-value is two times the lower tail area

Using \( t \) table: area in lower tail is between .05 and .10; therefore, \( p \)-value is between .10 and .20.

Using Excel: \( p \)-value = \( 2*T.DIST(-1.54,47,TRUE) = .1303 \)

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = \( .1304 \)

c. \( p \)-value > .05, do not reject \( H_0 \).

d. With \( df = 47 \), \( t_{.025} = 2.012 \)

Reject \( H_0 \) if \( t \leq -2.012 \) or \( t \geq 2.012 \)

\( t = -1.54 \); do not reject \( H_0 \)

26. a. \[
t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{103 - 100}{11.5 / \sqrt{65}} = 2.10
\]

Degrees of freedom = \( n - 1 = 64 \)

Because \( t > 0 \), \( p \)-value is two times the upper tail area
Using \( t \) table; area in upper tail is between .01 and .025; therefore, \( p \)-value is between .02 and .05.

Using Excel: \( p \)-value = 2*(1-T.DIST(2.10,64,TRUE)) = .0397

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0394

\( p \)-value \( \leq \) .05, reject \( H_0 \)

b. \[
\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{96.5 - 100}{11/\sqrt{65}} = -2.57
\]

Because \( t < 0 \), \( p \)-value is two times the lower tail area

Using \( t \) table: area in lower tail is between .005 and .01; therefore, \( p \)-value is between .01 and .02.

Using Excel: \( p \)-value = 2*T.DIST(-2.57,64,TRUE) = .0125

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0127

\( p \)-value \( \leq \) .05, reject \( H_0 \)

c. \[
\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{102 - 100}{10.5/\sqrt{65}} = 1.54
\]

Because \( t > 0 \), \( p \)-value is two times the upper tail area

Using \( t \) table: area in upper tail is between .05 and .10; therefore, \( p \)-value is between .10 and .20.

Using Excel: \( p \)-value = 2*(1-T.DIST(1.54,64,TRUE)) = .1285

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .1295

\( p \)-value > .05, do not reject \( H_0 \)

27. a. \( H_0: \mu \geq 13.04 \)

\( H_a: \mu < 13.04 \)  
Research hypothesis

b. \[
\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{12.75 - 13.04}{2/\sqrt{100}} = -1.45 \text{ (exact value)}
\]

Degrees of freedom = \( n - 1 = 99 \)

\( p \)-value is the lower tail area at the test statistic

Using \( t \) table: \( p \)-value is between .05 and .10

Using Excel: \( p \)-value = T.DIST(-1.45,99,TRUE) = .0751

c. \( p \)-value > .05; do not reject \( H_0 \). We cannot conclude that the cost of a restaurant meal is significantly cheaper than a comparable meal fixed at home.

d. \( df = 99 \quad t_{.05} = -1.66 \)

Reject \( H_0 \) if \( t \leq -1.66 \)

\(-1.45 > -1.66\); do not reject \( H_0 \)
28. a. $H_0: \mu \geq 9$
$H_1: \mu < 9$  
Challenge to the shareholders group claim

b. $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{7.27 - 9}{6.38/85} = -2.50$

Degrees of freedom $= n - 1 = 84$

$p$-value is lower-tail area

Using $t$ table: $p$-value is between .005 and .01
Using Excel: $p$-value = T.DIST(-2.50,84,TRUE) = .0072
Using unrounded Test Statistic via Excel with cell referencing, $p$-value = .0072

e. $p$-value $\leq .01$; reject $H_0$. The mean tenure of a CEO is significantly shorter than 9 years. The claim of the shareholders group is not valid.

30. a. $H_0: \mu = 6.4$
$H_1: \mu \neq 6.4$  
Research hypothesis

b. Using Excel and the datafile ChildCare, we find $\bar{x} = 7.0$ and $s = 2.4276$:

<table>
<thead>
<tr>
<th>Hours Spent on Child Care</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>

$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{7.0 - 6.4}{2.4276 / \sqrt{40}} = 1.56$

$df = n - 1 = 39$

Because $t > 0$, $p$-value is two times the upper tail area at $t = 1.56$

Using $t$ table: area in upper tail is between .05 and .10; therefore, $p$-value is between .10 and .20.
Using Excel: $p$-value = 2*(1-T.DIST(1.56,39,TRUE)) = .1268
Using unrounded Test Statistic calculated from the unrounded sample standard deviation via Excel with cell referencing, $p$-value = .1261
c. Most researchers would choose \( \alpha = .10 \) or less. If you chose \( \alpha = .10 \) or less, you cannot reject \( H_0 \). You are unable to conclude that the population mean number of hours married men with children in your area spend in child care differs from the mean reported by Time.

32. a. \( H_0: \mu = 10,192 \)
   \( H_1: \mu \neq 10,192 \) Research hypothesis

b. Using Excel and the datafile UsedCars, we find \( \bar{x} = 9750 \) and \( s = 1400 \):

<table>
<thead>
<tr>
<th>Sale Price</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>

\[
t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{9750 - 10,192}{1400 / \sqrt{50}} = -2.23
\]

Degrees of freedom = \( n - 1 = 49 \)

Because \( t < 0 \), \( p \)-value is two times the lower tail area

Using \( t \) table: area in lower tail is between .01 and .025; therefore, \( p \)-value is between .02 and .05.
Using Excel: \( p \)-value = 2*T.DIST(-2.23,49,TRUE) = .0304
Using unrounded Test Statistic calculated from the unrounded sample standard deviation via Excel with cell referencing, \( p \)-value = .0302

c. \( p \)-value \( \leq .05 \); reject \( H_0 \). The population mean price at this dealership differs from the national mean price $10,192.

34. a. \( H_0: \mu = 2 \)
   \( H_1: \mu \neq 2 \) Research hypothesis

b./c Inputting data given in the problem and using Excel, we find

<table>
<thead>
<tr>
<th>34 b, c</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>
</tbody>
</table>
Using formulas:

\[ \bar{x} = \frac{\sum x_i}{n} = \frac{22}{10} = 2.2 \]

c. \[ s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = .516 \]

d. \[ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{2.2 - 2}{.516/\sqrt{10}} = 1.22 \]

Degrees of freedom = \( n - 1 = 9 \)

Because \( t > 0 \), \( p \)-value is two times the upper tail area

Using \( t \) table: area in upper tail is between .10 and .20; therefore, \( p \)-value is between .20 and .40.

Using Excel: \( p \)-value = \( 2 \times (1 - \text{T.DIST}(1.22,9,\text{TRUE})) = .2535 \)

Using unrounded Test Statistic calculated from the unrounded sample standard deviation via Excel with cell referencing, \( p \)-value = .2518

e. \( p \)-value > .05; do not reject \( H_0 \). No reason to change from the 2 hours for cost estimating purposes.

36. a. \[ z = \frac{\bar{p} - p_o}{\sqrt{p_o(1-p_o)/n}} = \frac{.68 - .75}{\sqrt{.75(1-.75)/300}} = \frac{-2.80}{\text{exact value}} \]

\( p \)-value is lower-tail area

Using normal table with \( z = -2.80 \): \( p \)-value = .0026

Using Excel: \( p \)-value = \( \text{NORM.S.DIST}(-2.80,\text{TRUE}) = .0026 \)

\( p \)-value ≤ .05; reject \( H_0 \). The proportion is less than .75.

b. \[ z = \frac{.72 - .75}{\sqrt{.75(1-.75)/300}} = -1.20 \] (exact value)

\( p \)-value is lower-tail area

Using normal table with \( z = -1.20 \): \( p \)-value = .1151

Using Excel: \( p \)-value = \( \text{NORM.S.DIST}(-1.20,\text{TRUE}) = .1151 \)

\( p \)-value > .05; do not reject \( H_0 \). We cannot conclude that the proportion is less than .75.
c. \[ z = \frac{.70 - .75}{\sqrt{\frac{.75(1-.75)}{300}}} = -2.00 \] (exact value)

\[ p \text{-value is lower-tail area} \]

Using normal table with \( z = -2.00 \): \( p \text{-value} = .0228 \)
Using Excel: \( p \text{-value} = \text{NORM.S.DIST}(-2.00,\text{TRUE}) = .0228 \)
\( p \text{-value} \leq .05; \) reject \( H_0 \). The proportion is less than .75.

d. \[ z = \frac{.77 - .75}{\sqrt{\frac{.75(1-.75)}{300}}} = .80 \] (exact value)

\[ p \text{-value is lower-tail area} \]

Using normal table with \( z = .80 \): \( p \text{-value} = .7881 \)
Using Excel: \( p \text{-value} = \text{NORM.S.DIST}(.80,\text{TRUE}) = .7881 \)
\( p \text{-value} > .05; \) do not reject \( H_0 \). We cannot conclude that the proportion is less than .75.

38. a. \( H_0: p = .64 \)
\( H_a: p \neq .64 \) Research hypothesis

b. \( \bar{p} = \frac{52}{100} = .52 \)

\[ z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.52 - .64}{\sqrt{\frac{.64(1-.64)}{100}}} = -2.50 \] (exact value)

Because \( z < 0 \), \( p \text{-value} \) is two times the lower tail area

Using normal table with \( z = -2.50 \): \( p \text{-value} = 2(.0062) = .0124 \)
Using Excel: \( p \text{-value} = 2*\text{NORM.S.DIST}(-2.50,\text{TRUE}) = .0124 \)

c. \( p \text{-value} \leq .05; \) reject \( H_0 \). Proportion differs from the reported .64.

d. Yes. Since \( \bar{p} = .52 \), it indicates that fewer than 64\% of the shoppers believe the supermarket brand is as good as the name brand.

40. a. Sample proportion: \( \bar{p} = .35 \)

Number planning to provide holiday gifts: \( n\bar{p} = 60(.35) = 21 \)

b. \( H_0: p \geq .46 \)
\( H_a: p < .46 \) Research hypothesis

\[ z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.35 - .46}{\sqrt{\frac{.46(1-.46)}{60}}} = -1.71 \]

\( p \text{-value is area in lower tail} \)
42. a. \( \bar{p} = \frac{12}{80} = .15 \)

b. \[ \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.15(.85)}{80}} = .0399 \]

\[ \bar{p} \pm z_{.025} \sqrt{\frac{p(1-p)}{n}} \]

\[ .15 \pm 1.96(.0399) \]

\[ .15 \pm .0782 \text{ or } .0718 \text{ to } .2282 \]

c. We can conduct a hypothesis test concerning whether the return rate for the Houston store is equal to .06 at an \( \alpha = .05 \) level of significance using the 95% confidence interval in part (b). Since the confidence interval does not include .06, we conclude that the return rate for the Houston store is different than the U.S. national return rate.

44. a. \( H_0: p \leq .50 \quad H_a: p > .50 \quad \text{Research hypothesis} \)

b. Using Excel Lawsuit data file, we find that 92 of the 150 physicians in the sample have been sued.

So, \( \bar{p} = \frac{92}{150} = .6133 \)

\[ z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.6133-.50}{\sqrt{\frac{(.50)(.50)}{150}}} = 2.78 \]

\( p\)-value is the area in the upper tail at \( z = 2.78 \)

Using normal table with \( z = 2.78 \): \( p\)-value = 1 - .9973 = .0027

Using Excel: \( p\)-value = 1 - NORM.S.DIST(2.78,TRUE) = .0027

Using unrounded proportion and Test Statistic via Excel with cell referencing, \( p\)-value = .0028

c. Since \( p\)-value = .0027 \( \leq \) .01, we reject \( H_0 \) and conclude that the proportion of physicians over the age of 55 who have been sued at least once is greater than .50.
46. \( H_0: \mu = 101.5 \)
   \( H_a: \mu \neq 101.5 \)  
   **Research hypothesis**

Using the FedEmail datafile and Excel, the sample mean, \( \bar{x} = 100.47 \)

\[
   z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{100.47 - 101.5}{25/\sqrt{10163}} = -4.15
\]

Because \( z < 0 \), \( p \)-value is two times the lower tail area

Using normal table with \( z = -4.15 \): \( p \)-value \( \approx 2(.0000) = .0000 \)
Using Excel: \( p \)-value = \( 2 \times \text{NORM.S.DIST}(-4.15, \text{TRUE}) \) = .0000
Using the unrounded sample mean to calculate the Test Statistic via Excel with cell referencing, \( p \)-value = .0000

\( p \)-value \( \leq .01 \), reject \( H_0 \). Conclude that the actual mean number of business emails sent and received per business day by employees of this department of the Federal Government differs from corporate employees.

Although the difference between the sample mean number of business emails sent and received per business day by employees of this department of the Federal Government and the mean number of business emails sent and received by corporate employees is statistically significant, this difference of 1.03 emails is relatively small and so may be of little or no practical significance.

48. \( H_0: p \leq .62 \)
   \( H_a: p > .62 \)  
   **Research hypothesis**

\[
   \bar{p} = \frac{31038}{49581} = .626
\]

\[
   z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.626 - .62}{\sqrt{\frac{.62(1-.62)}{49581}}} = 2.75
\]

\( p \)-value is the area in the upper tail

Using normal table with \( z = 2.75 \): \( p \)-value = \( 1 - .9970 = .0030 \)
Using Excel: \( p \)-value = \( 1 - \text{NORM.S.DIST}(2.75, \text{TRUE}) \) = .0030
Using unrounded proportion and Test Statistic via Excel with cell referencing, \( p \)-value = .0029

\( p \)-value \( \leq .05 \), reject \( H_0 \). Conclude that the actual proportion of fast food orders this year that includes French fries exceeds the proportion of fast food orders that included French fries last year.

Although the difference between the sample proportion of fast food orders this year that includes French fries and the proportion of fast food orders that included French fries last year is statistically significant, APGA should be concerned about whether this .006 or .6% increase is large enough to be effective in an advertising campaign.

50. a. \( H_0: \mu = 16 \)
    \( H_a: \mu \neq 16 \)  
    **Research hypothesis to test for over or under filling**

\[
   z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{16.32 - 16}{.8/\sqrt{30}} = 2.19
\]

Because \( z > 0 \), \( p \)-value is two times the upper tail area

Using normal table with \( z = 2.19 \): \( p \)-value = \( 2(0.0143) = .0286 \)
Hypothesis Tests

Using Excel: \( p\)-value = \(2 \times (1 - \text{NORM.S.DIST}(2.19, \text{TRUE})) = .0285 \)
Using unrounded Test Statistic via Excel with cell referencing, \( p\)-value = .0285

\( p\)-value \leq .05; reject \( H_0 \). Readjust production line.

c. \[ z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{15.82 - 16}{.8 / \sqrt{30}} = -1.23 \]

Because \( z < 0 \), \( p\)-value is two times the lower tail area

Using normal table with \( z = -1.23 \): \( p\)-value = \(2 \times (1.093) = .2186 \)
Using Excel: \( p\)-value = \(2 \times \text{NORM.S.DIST}(-1.23, \text{TRUE}) = .2187 \)
Using unrounded Test Statistic via Excel with cell referencing, \( p\)-value = .2178

\( p\)-value > .05; do not reject \( H_0 \). Continue the production line.

d. Reject \( H_0 \) if \( z \leq -1.96 \) or \( z \geq 1.96 \)

For \( \bar{x} = 16.32 \), \( z = 2.19 \); reject \( H_0 \)

For \( \bar{x} = 15.82 \), \( z = -1.23 \); do not reject \( H_0 \)

Yes, same conclusion.

52. a. \( H_0 \): \( \mu \leq 4 \)
\( H_a \): \( \mu > 4 \) Research hypothesis

b. \[ z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{4.5 - 4}{1.5 / \sqrt{80}} = 2.58 \]

\( p\)-value is the upper tail area at \( z = 2.58 \)

Using normal table with \( z = 2.58 \): \( p\)-value = \(1.0000 - .9951 = .0049 \)
Using Excel: \( p\)-value = \(1 - \text{NORM.S.DIST}(2.58, \text{TRUE}) = .0049 \)
Using unrounded Test Statistic via Excel with cell referencing, \( p\)-value = .0049

c. \( p\)-value \leq .01, reject \( H_0 \). Conclude that the mean daily background television that children from low-income families are exposed to is greater than four hours.

54. \( H_0 \): \( \mu \leq 30.8 \)
\( H_a \): \( \mu > 30.8 \) Research hypothesis

Using Excel and the datafile BritainMarriages, we find \( \bar{x} = 32.72 \) and \( s = 12.10 \):

| Age |
|-----------------
| Mean 32.72340426 |
| Standard Error 1.76556174 |
| Median 31 |
| Mode 21 |
| Standard Deviation 12.10408147 |
| Sample Variance 146.5087882 |
| Kurtosis 2.118162618 |
Chapter 9

Skewness 1.282462905
Range 56
Minimum 18
Maximum 74
Sum 1538
Count 47

\[
\bar{x} = 32.72 \\
\sigma = 12.10 \\
\frac{t}{s / \sqrt{n}} = \frac{32.72 - 30.8}{12.10 / \sqrt{47}} = 1.09 \\
\text{Degrees of freedom} = 47 - 1 = 46 \\
p\text{-value is the upper tail area} \\
\text{Using } t\text{ table: area in upper tail is between .10 and .20; therefore, } p\text{-value is between .10 and .20.}
\]

Using Excel: \( p\text{-value} = 1 - T.DIST(1.09, 46, \text{TRUE}) = 0.1407 \)

Using unrounded Test Statistic calculated from the unrounded sample standard deviation via Excel with cell referencing, \( p\text{-value} = .1408 \)

\( p\text{-value} > .05; \) do not reject \( H_0. \)

The mean age of British men at the time of marriage exceeds the 2013 mean age of 30.8.

56. \( H_0: \mu \leq 125,000 \) \hspace{1cm} Chamber of Commerce claim
\( H_a: \mu > 125,000 \)

\[
\frac{t}{s / \sqrt{n}} = \frac{130,000 - 125,000}{12,500 / \sqrt{32}} = 2.26 \\
\text{Degrees of freedom} = 32 - 1 = 31 \\
p\text{-value is upper-tail area} \\
\text{Using } t\text{ table: } p\text{-value is between .01 and .025} \\
\text{Using Excel: } p\text{-value} = 1 - T.DIST(2.26, 31, \text{TRUE}) = .0155 \\
\text{Using unrounded Test Statistic via Excel with cell referencing, } p\text{-value} = .0154 \\
\]

\( p\text{-value} \leq .05; \) reject \( H_0. \) Conclude that the mean cost is greater than $125,000 per lot.

58. a. \( H_0: p \leq .52 \)  
\( H_a: p > .52 \) \hspace{1cm} Research hypothesis

\[
\bar{p} = \frac{285}{510} = .5588 \\
\frac{z}{n} = \frac{.5588 - .52}{.52(1 - .52)} / \sqrt{510} = 1.75 \\
p\text{-value is the area in the upper tail} \\
\text{Using normal table with } z = 1.75; \text{ } p\text{-value} = 1.0000 - .9599 = .0401
Using Excel: \( p\)-value = 1-NORM.S.DIST(1.75,TRUE) = .0401
Using unrounded proportion and Test Statistic via Excel with cell referencing, \( p\)-value = .0396

\( p\)-value \(\leq\) .05; reject \( H_0 \). We conclude that people who fly frequently are more likely to be able to sleep during flights.

b. \( H_0: p \leq .52 \)
\( H_a: p > .52 \)

\[
\bar{p} = \frac{285}{510} = .5588
\]

\[
z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.5588 - .52}{\sqrt{\frac{.52(1-.52)}{510}}} = 1.75
\]

\( p\)-value is the area in the upper tail

Using normal table with \( z = 1.75 \); \( p\)-value = 1.0000 - .9599 = .0401
Using Excel: \( p\)-value = 1-NORM.S.DIST(1.75,TRUE) = .0401
Using unrounded proportion and Test Statistic via Excel with cell referencing, \( p\)-value = .0396

\( p\)-value > .01; we cannot reject \( H_0 \). Thus, we cannot conclude that that people who fly frequently better more likely to be able to sleep during flights.

60. a. \( H_0: p \leq .30 \) Organization claim
\( H_a: p > .30 \) Research hypothesis

b. \( \bar{p} = \frac{136}{400} = .34 \) (34%)

c. \[
z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.34 - .30}{\sqrt{\frac{.30(1-.30)}{400}}} = 1.75
\]

\( p\)-value is the upper-tail area

Using normal table with \( z = 1.75 \); \( p\)-value = 1.0000 - .9599 = .0401
Using Excel: \( p\)-value = 1-NORM.S.DIST(1.75,TRUE) = .0401
Using unrounded Test Statistic via Excel with cell referencing, \( p\)-value = .0404

d. \( p\)-value \(\leq\) .05; reject \( H_0 \). Conclude that more than 30\% of the millennials either live at home with their parents or are otherwise dependent on their parents.

62. \( H_0: p \geq .90 \) Radio station claim
\( H_a: p < .90 \) Research hypothesis

\[
\bar{p} = \frac{49}{58} = .8448
\]

\[
z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.8448 - .90}{\sqrt{\frac{.90(1-.90)}{58}}} = -1.40
\]

\( p\)-value is the lower-tail area
Using normal table with $z = -1.40$: $p$-value = .0808
Using Excel: $p$-value = NORM.S.DIST(-1.40,TRUE) = .0808
Using unrounded proportion and Test Statistic via Excel with cell referencing, p-value = .0807

$p$-value > .05; do not reject $H_0$. Claim of at least 90% cannot be rejected.
Chapter 10
Statistical Inference About Means and Proportions with Two Populations

Solutions:

1. a. \( \bar{x}_1 - \bar{x}_2 = 13.6 - 11.6 = 2 \)
   
b. \( z_{a/2} = z_{.05} = 1.645 \)

   \[
   \bar{x}_1 - \bar{x}_2 \pm 1.645 \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}
   \]

   \[
   2 \pm 1.645 \sqrt{\frac{(2.2)^2}{50} + \frac{(3)^2}{35}}
   \]

   \[ 2 \pm .98 \quad (1.02 \text{ to } 2.98) \]

c. \( z_{a/2} = z_{.025} = 1.96 \)

   \[
   2 \pm 1.96 \sqrt{\frac{(2.2)^2}{50} + \frac{(3)^2}{35}}
   \]

   \[ 2 \pm 1.17 \quad (.83 \text{ to } 3.17) \]

2. a. \( z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}} = \frac{(25.2 - 22.8) - 0}{\sqrt{\frac{(5.2)^2}{40} + \frac{6^2}{50}}} = 2.03 \)

   b. \( p\)-value is upper-tail area

   Using normal table with \( z = 2.03 \): \( p\)-value = 1.0000 - .9788 = .0212
   Using Excel: \( p\)-value = 1-NORM.S.DIST(2.03,TRUE) = .0212
   Using unrounded Test Statistic via Excel with cell referencing, \( p\)-value = .0211

c. \( p\)-value \( \leq .05 \), reject \( H_0 \).

4. a. \( \mu_1 \) = population mean for smaller cruise ships
   \( \mu_2 \) = population mean for larger cruise ships

   \[ \bar{x}_1 - \bar{x}_2 = 85.36 - 81.40 = 3.96 \]
b. 

\[ z_{0.05} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \]

\[ 1.96 \sqrt{\frac{(4.55)^2}{37} + \frac{(3.97)^2}{44}} = 1.88 \]

c. 3.96±1.88 (2.08 to 5.84)

6. \( \mu_1 = \) mean hotel price in Atlanta  
\( \mu_2 = \) mean hotel price in Houston

\( H_0: \mu_1 - \mu_2 \geq 0 \)  
\( H_a: \mu_1 - \mu_2 < 0 \) Research hypothesis

Using the Hotel data file and Excel's Data Analysis Tool, the Two Sample z-Test results are:

<table>
<thead>
<tr>
<th></th>
<th>Atlanta</th>
<th>Houston</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>91.71429</td>
<td>101.125</td>
</tr>
<tr>
<td>Known Variance</td>
<td>400</td>
<td>625</td>
</tr>
<tr>
<td>Observations</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( z )</td>
<td>-1.8093</td>
<td></td>
</tr>
<tr>
<td>P(Z&lt;=z) one-tail</td>
<td>0.035202</td>
<td></td>
</tr>
<tr>
<td>z Critical one-tail</td>
<td>1.644854</td>
<td></td>
</tr>
<tr>
<td>P(Z&lt;=z) two-tail</td>
<td>0.070405</td>
<td></td>
</tr>
<tr>
<td>z Critical two-tail</td>
<td>1.959964</td>
<td></td>
</tr>
</tbody>
</table>

\[ z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(91.71-101.13) - 0}{\sqrt{\frac{20^2}{35} + \frac{25^2}{40}}} = -1.81 \]

\( p \)-value is the lower-tail area

Using normal table with \( z = -1.81 \): \( p \)-value = .0351  
Using Excel: \( p \)-value = NORM.S.DIST(-1.81,TRUE) = .0351  
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0352  
\( p \)-value \( \leq .05 \); reject \( H_0 \). The mean price of a hotel room in Atlanta is lower than the mean price of a hotel room in Houston.

8. a. \( \mu_1 = \) population mean Year 2 score  
\( \mu_2 = \) population mean Year 1 score

\( H_0: \mu_1 - \mu_2 \leq 0 \)  
\( H_a: \mu_1 - \mu_2 > 0 \) Research hypothesis
Statistical Inference About Means and Proportions with Two Populations

\[
z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(76 - 73)}{\sqrt{\frac{6^2}{60} + \frac{6^2}{60}}} = 2.74
\]

\(p\)-value is upper-tail area

Using normal table with \(z = 2.74\): \(p\)-value = 1.0000 - .9969 = .0031  
Using Excel: \(p\)-value = 1-NORM.S.DIST(2.74,TRUE) = .0031  
Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .0031

\(p\)-value \(\leq .05\), we reject the null hypothesis. The difference is significant. We can conclude that customer service has improved for Rite Aid.

b. This is another upper tail test but it only involves one population.

\(H_0: \mu \leq 75.7\)  
\(H_a: \mu > 75.7\)  
Research hypothesis

\[
z = \frac{\bar{x}_1 - \mu_0}{\sigma / \sqrt{n}} = \frac{76 - 75.7}{6 / \sqrt{60}} = .39
\]

\(p\)-value is upper-tail area

Using normal table with \(z = .39\): \(p\)-value = 1.0000 - .6517 = .3483  
Using Excel: \(p\)-value = 1-NORM.S.DIST(.39,TRUE) = .3483  
Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .3493

\(p\)-value > .05, we cannot reject the null hypothesis. The difference is not statistically significant.

c. This is an upper tail test similar to the one in part (a).

\(H_0: \mu_1 - \mu_2 \leq 0\)  
\(H_a: \mu_1 - \mu_2 > 0\)  
Research hypothesis

\[
z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(77 - 75)}{\sqrt{\frac{6^2}{60} + \frac{6^2}{60}}} = 1.83
\]

\(p\)-value is upper-tail area

Using normal table with \(z = 1.83\): \(p\)-value = 1.0000 - .9664 = .0336  
Using Excel: \(p\)-value = 1-NORM.S.DIST(1.83,TRUE) = .0336  
Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .0339

\(p\)-value \(\leq .05\), we reject the null hypothesis. The difference is significant. We can conclude that customer service has improved for Expedia.

d. We will reject the null hypothesis of “no increase” if the \(p\)-value \(\leq .05\). For an upper tail hypothesis test, the \(p\)-value is the area in the upper tail at the value of the test statistic. A value of \(z = 1.645\) provides an upper tail area of .05. So, we must solve the following equation for \(\bar{x}_1 - \bar{x}_2\). 

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
Chapter 10

\[ z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 1.645 \]

\[ \bar{x}_1 - \bar{x}_2 = 1.645\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.80 \]

This tells us that as long as the Year 2 score for a company exceeds the Year 1 score by more than 1.80, the difference will be statistically significant.

e. The increase from Year 1 to Year 2 for J.C. Penney is not statistically significant because it is less than 1.80. We cannot conclude that customer service has improved for J.C. Penney.

9. a. \( \bar{x}_1 - \bar{x}_2 = 22.5 - 20.1 = 2.4 \)

b. \[ df = \frac{1}{\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}} = \frac{1}{\frac{1}{20} + \frac{1}{30}} = 45.8 \]

Use \( df = 45 \).

e. \( t_{0.025} = 2.014 \)

\[ t_{0.025} = \frac{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}{\bar{x}_1 - \bar{x}_2} = 2.014 \sqrt{\frac{2.5^2}{20} + \frac{4.8^2}{30}} = 2.1 \]

d. \( 2.4 \pm 2.1\) (.3 to 4.5)

10. a. \( t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5.2^2}{35} + \frac{8.5^2}{40} = 2.18 \)

b. \[ df = \frac{1}{\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}} = \frac{1}{\frac{1}{34} + \frac{1}{39}} = 65.7 \]

c. Degrees of freedom = 65

Because \( t > 0 \), \( p \)-value is two times the upper tail area

Using \( t \) table; area in upper tail is between .01 and .025; therefore, \( p \)-value is between .02 and .05.

Using Excel: \( p \)-value = 2*(1-T.DIST(2.18,65,TRUE)) = .0329

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0329

d. \( p \)-value \( \leq .05 \), reject \( H_0 \).
12. a. \( \mu_1 = \) population mean miles that Buffalo residents travel per day  
\( \mu_2 = \) population mean miles that Boston residents travel per day  
\[ \bar{x}_1 - \bar{x}_2 = 22.5 - 18.6 = 3.9 \]

b. \[
\begin{align*}
\text{df} & = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1-1} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2-1} \left( \frac{s_2^2}{n_2} \right)^2} \\
& = \frac{\left( \frac{8.4^2}{50} + \frac{7.4^2}{40} \right)^2}{1} \\
& = \left( \frac{8.4^2}{50} + \frac{7.4^2}{40} \right)^2 = 87.1
\end{align*}
\]

Use \( df = 87, \ t_{0.025} = 1.988 \)

\[ 3.9 \pm 1.988 \sqrt{\frac{8.4^2}{50} + \frac{7.4^2}{40}} \]

\[ 3.9 \pm 3.3 \] (.6 to 7.2)

14. a. \( H_0: \mu_1 - \mu_2 \geq 0 \)
\( H_a: \mu_1 - \mu_2 < 0 \) Research hypothesis

b. \[
\begin{align*}
\text{t} & = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\
& = \frac{(56,100 - 59,400) - 0}{\sqrt{40 + 50}} \\
& = -2.41
\end{align*}
\]

c. \[
\begin{align*}
\text{df} & = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1-1} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2-1} \left( \frac{s_2^2}{n_2} \right)^2} \\
& = \frac{\left( \frac{4000^2}{40} + \frac{7000^2}{50} \right)^2}{1} \\
& = \left( \frac{4000^2}{40} + \frac{7000^2}{50} \right)^2 = 87.55
\end{align*}
\]

Degrees of freedom = 87  
\( p \)-value is lower-tail area

Using \( t \) table: \( p \)-value is between .005 and .01
Using Excel: \( p \)-value = T.DIST(-2.41,87,TRUE) = .0090
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0091

d. \( p \)-value \( \leq .05 \), reject \( H_0 \). We conclude that the salaries of staff nurses are lower in Tampa than in Dallas.

16. a. \( \mu_1 = \) population mean math score parents college grads  
\( \mu_2 = \) population mean math score parents high school grads  
\( H_0: \mu_1 - \mu_2 \leq 0 \)  
\( H_a: \mu_1 - \mu_2 > 0 \) Research hypothesis

b. Using the SATMath data file and Excel’s Data Analysis Tool, the Two Sample t-Test with Unequal Variances results are:
t-Test: Two-Sample Assuming Unequal Variances

<table>
<thead>
<tr>
<th></th>
<th>College</th>
<th>High School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>525</td>
<td>487</td>
</tr>
<tr>
<td>Variance</td>
<td>3530.8</td>
<td>2677.818182</td>
</tr>
<tr>
<td>Observations</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>1.80375262</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) one-tail</td>
<td>0.04166737</td>
<td></td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1.70814076</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.08333474</td>
<td></td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.05953855</td>
<td></td>
</tr>
</tbody>
</table>

\[ \bar{x}_1 = \frac{\sum x_i}{n} = \frac{8400}{16} = 525 \]

\[ \bar{x}_2 = \frac{\sum x_i}{n} = \frac{5844}{12} = 487 \]

\[ \bar{x}_1 - \bar{x}_2 = 525 - 487 = 38 \text{ points higher if parents are college grads} \]

c. \[ s_1 = \sqrt{\frac{\sum (x_i - \bar{x}_1)^2}{n_1-1}} = \sqrt{\frac{52962}{16-1}} = \sqrt{3530.8} = 59.42 \]

\[ s_2 = \sqrt{\frac{\sum (x_i - \bar{x}_2)^2}{n_2-1}} = \sqrt{\frac{29456}{12-1}} = \sqrt{2677.82} = 51.75 \]

\[ t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s_1 \sqrt{n_1} + s_2 \sqrt{n_2}} = \frac{(525 - 487) - 0}{59.42 \sqrt{16} + 51.75 \sqrt{12}} = 1.80 \]

\[ df = \frac{1}{n_1-1} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2-1} \left( \frac{s_2^2}{n_2} \right)^2 = \frac{1}{15} \left( \frac{59.42^2}{16} \right) + \frac{1}{11} \left( \frac{51.75^2}{12} \right) = 25.3 \]

Degrees of freedom = 25

\( p \)-value is upper-tail area

Using \( t \) table: \( p \)-value is between .025 and .05

Using Excel: \( p \)-value = 1-\( \text{T.DIST(1.80,25,TRUE)} \) = .0420

Using unrounded standard deviations and the resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0417

d. \( p \)-value \( \leq .05 \), reject \( H_0 \). Conclude higher population mean math scores for students whose parents are college grads.
18. a. Let \( \mu_1 \) = population mean minutes late for delayed AirTran flights
\( \mu_2 \) = population mean minutes late for delayed Southwest flights

\( H_0: \mu_1 - \mu_2 = 0 \)
\( H_a: \mu_1 - \mu_2 \neq 0 \)  
Research hypothesis

b. Using the AirDelay data file and Excel’s Data Analysis Tool, the Two Sample t-Test with Unequal Variances results are:

<table>
<thead>
<tr>
<th>Test</th>
<th>AirTran</th>
<th>Southwest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>50.6</td>
<td>52.8</td>
</tr>
<tr>
<td>Variance</td>
<td>705.75</td>
<td>404.378947</td>
</tr>
<tr>
<td>Observations</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>-0.316067985</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) one-tail</td>
<td>0.376740031</td>
<td></td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1.681070703</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.753480062</td>
<td></td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.016692199</td>
<td></td>
</tr>
</tbody>
</table>

The difference between sample mean delay times is 50.6 – 52.8 = -2.2 minutes, which indicates the sample mean delay time is 2.2 minutes less for AirTran Airways.

c. Sample standard deviations: \( s_1 = 26.57 \) and \( s_2 = 20.11 \)

\[
t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{(50.6 - 52.8) - 0}{\sqrt{26.57^2/25 + 20.11^2/20}} = -0.32
\]

\[
df = \frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2 = \frac{1}{24}\left(\frac{26.57^2}{25}\right)^2 + \frac{1}{19}\left(\frac{20.11^2}{20}\right)^2 = 42.9
\]

Degrees of freedom = 42
Because \( t < 0 \), p-value is two times the lower tail area
Using $t$ table: area in lower tail is greater than .20; therefore, $p$-value is greater than .40.
Using Excel: $p$-value $= 2 \times T.DIST(-.32,42,TRUE) = .7506$
Using unrounded standard deviations and the resulting unrounded Test Statistic via Excel with cell referencing, $p$-value $= .7535$

$p$-value $>.05$, do not reject $H_0$. We cannot reject the assumption that the population mean delay times are the same at AirTran Airways and Southwest Airlines. There is no statistical evidence that one airline does better than the other in terms of their population mean delay time.

19. a. 1, 2, 0, 0, 2
b. $\bar{d} = \frac{\sum d_i}{n} = \frac{5}{5} = 1$
c. $s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{4}{5-1}} = 1$
d. $t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{1-0}{1/\sqrt{5}} = 2.24$

Degrees of freedom $= n - 1 = 4$
$p$-value is upper-tail area

Using $t$ table: $p$-value is between .025 and .05
Using Excel: $p$-value $= 1 - T.DIST(2.24,4,TRUE) = .0443$
Using unrounded Test Statistic via Excel with cell referencing, $p$-value $= .0445$

$p$-value $\leq .05$, Reject $H_0$; conclude $\mu_d > 0$.

20. a. 3, -1, 3, 5, 3, 0, 1
b. $\bar{d} = \frac{\sum d_i}{n} = \frac{14}{7} = 2$
c. $s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{26}{7-1}} = 2.082$
d. $\bar{d} = 2$
e. With 6 degrees of freedom $t_{0.025} = 2.447$

$2 \pm 2.447 \left( \frac{2.082}{\sqrt{7}} \right)$

$2 \pm 1.93 \quad (.07 \text{ to } 3.93)$

21. $x_i$ = After Rating score
$x_i$ = Before Rating score

Difference = rating after - rating before

$H_0$: $\mu_d \leq 0$
$H_a$: $\mu_d > 0$ Research hypothesis
\[ \bar{d} = .625 \text{ and } s_d = 1.30 \]

\[ t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{.625 - 0}{1.30 / \sqrt{8}} = 1.36 \]

Degrees of freedom = \( n - 1 = 7 \)

\( p \)-value is upper-tail area

Using \( t \) table: \( p \)-value is between .10 and .20
Using Excel: \( p \)-value = 1-\( \text{T.DIST}(1.36,7,\text{TRUE}) \) = .1080
Using unrounded standard deviation and the resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .1084

\( p \)-value >.05, Do not reject \( H_0 \); we cannot conclude that seeing the commercial improves the mean potential to purchase.

22. a. \( x_i \) = 1st quarter price per share
\( x_i \) = Beginning of year price per share

Let \( d_i \) = 1st quarter price per share – beginning of year price per share

Using the StockPrices data file and Excel’s Data Analysis Tool, Descriptive Statistics results of the differences are:

<table>
<thead>
<tr>
<th>Stock Price Difference (d) 1st Qtr - Beg of Yr</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>
<tr>
<td>Confidence Level(95.0%)</td>
</tr>
</tbody>
</table>

\[ \bar{d} = \frac{\sum d_i}{n} = \frac{85.25}{25} = 3.41 \]

\[ s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{428.26}{25-1}} = 4.22 \]

With \( df = 24, t_{.025} = 2.064 \)

\[ \bar{d} \pm t_{.025} \frac{s_d}{\sqrt{n}} \]
Confidence interval: $3.41 \pm 1.74 \quad ($1.67 to $5.15$)

The 95\% confidence interval shows that the population mean price per share of stock has increased between $1.67 and $5.15 over the three-month period.

Note that at the beginning of year

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1153.16}{25} = 46.13$$

With this as the sample mean price per share of stock at the beginning of 2012, the confidence interval ($1.67 to $5.15) indicates the percentage change in the population mean price per share of stock would have increased from

$$1.67/46.13 = .036, \text{ or } 3.6\%$$
$$to \quad 5.15/46.13 = .112, \text{ or } 11.2\%$$

Thus, for the population of stocks, the mean price per share has increased between 3.6\% and 11.2\% over the three-month period. This was excellent news for the 1st quarter of 2012. Stock prices were having one of the largest quarterly increases in years. The outlook for a recovering economy was very good at the end of the 1st quarter of 2012.

24. a. \( x_1 = \text{Current Year Airfare} \)
\( x_2 = \text{Previous Year Airfare} \)
Difference = Current Year Airfare – Previous Year Airfare

\( H_0: \mu_d \leq 0 \)
\( H_a: \mu_d > 0 \) 
Research hypothesis

Using the BusinessTravel data file and Excel’s Data Analysis Tool, Descriptive Statistics and Paired t test results are:

<table>
<thead>
<tr>
<th>d</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>Mean 23</td>
</tr>
<tr>
<td>-42</td>
<td>Standard Error 11.199973</td>
</tr>
<tr>
<td>10</td>
<td>Median 30</td>
</tr>
<tr>
<td>10</td>
<td>Mode 30</td>
</tr>
<tr>
<td>-27</td>
<td>Standard Dev 38.797844</td>
</tr>
<tr>
<td>50</td>
<td>Sample Var 1505.2727</td>
</tr>
<tr>
<td>60</td>
<td>Kurtosis -1.167691</td>
</tr>
<tr>
<td>60</td>
<td>Skewness -0.575146</td>
</tr>
<tr>
<td>-30</td>
<td>Range 105</td>
</tr>
<tr>
<td>62</td>
<td>Minimum -42</td>
</tr>
<tr>
<td>30</td>
<td>Maximum 63</td>
</tr>
<tr>
<td>Sum 276</td>
<td>Count 12</td>
</tr>
</tbody>
</table>

| t-Test: Paired Two Sample for Means |
|---|---|---|
| Current Year | Previous Year |
| Mean | 487 | 464 |
| Variance | 23238 | 18508.54545 |
| Hypothesized Mean Diff | 0 |
| df | 11 |
| t Stat | 2.053576389 |
| P(T<=t) one-tail | 0.032288261 |
| t Critical one-tail | 1.79588419 |
| P(T<=t) two-tail | 0.064576523 |
| t Critical two-tail | 2.20098516 |
Differences $30, 63, -42, 10, 10, -27, 50, 60, 60, -30, 62, 30$

\[
\bar{d} = \frac{\Sigma d_i}{n} = \frac{216}{12} = 23
\]

\[
s_d = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{16.558}{12-1}} = 38.80
\]

\[
t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{23 - 0}{38.80 / \sqrt{12}} = 2.05
\]

Degrees of freedom $= n - 1 = 11$

$p$-value is upper-tail area

Using $t$ table: $p$-value is between .025 and .05

Using Excel: $p$-value = $1 - \text{T.DIST}(2.05, 11, \text{TRUE}) = .325$

Using unrounded standard deviation and resulting unrounded Test Statistic via Excel with cell referencing, $p$-value = .0323

Since $p$-value < .05, reject $H_0$. We can conclude that there has been a significance increase in business travel airfares over the one-year period.

b. Current year: $\bar{x} = \frac{\Sigma x_i}{n} = \frac{5844}{12} = $487

Previous year: $\bar{x} = \frac{\Sigma x_i}{n} = \frac{5568}{12} = $464

c. One-year increase = $487 - $464 = $23

$23/464 = .05$, or a 5% increase in business travel airfares for the one-year period.

26. a. $x_1$ = First round score

$x_2$ = Final round score

Difference = First round – Final Round

$H_0: \mu_d = 0$

$H_a: \mu_d \neq 0$  

Research hypothesis

Using the GolfScores data file and Excel’s Data Analysis Tool, Descriptive Statistics and Paired $t$ test results are:

d | Difference
--- | ---
-2 | 
-1 | 
-5 | 0.74153113
1 | 
1 | 
0 | 3.31622804
4 | 10.9973684
-7 | -0.7763571
-6 | -0.547846

<p>| t-Test: Paired Two Sample for Means |
| --- | --- |
| <strong>Round 1</strong> | <strong>Round 2</strong> |
| Mean | 69.65 | 70.7 |
| Variance | 2.7657895 | 9.168421 |
| Observations | 20 | 20 |
| Hypothesized Mean Diff | 0 |
| df | 19 |
| t Stat | -1.415989 |</p>
<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th></th>
<th>P(T&lt;=t) one-tail</th>
<th></th>
<th>t Critical one-tail</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Minimum</td>
<td>-7</td>
<td>t Critical one-tail</td>
<td>1.3277282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Maximum</td>
<td>4</td>
<td>P(T&lt;=t) two-tail</td>
<td>0.1729638</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>Sum</td>
<td>-21</td>
<td>t Critical two-tail</td>
<td>1.7291328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-7</td>
<td>Count</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Differences:  -2, -1, -5, 1, 1, 0, 4, -7, -6, 1, 0, 2, -3, -7, -2, 3, 1, 2, 1, -4

\[
d = \frac{\sum d_i}{n} = \frac{-21}{20} = -1.05
\]

\[
s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = 3.3162
\]

\[
t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{-1.05 - 0}{3.3162 / \sqrt{20}} = -1.42
\]

Degrees of freedom = \( n - 1 = 19 \)

Because t<0, p-value is two times the lower tail area

Using \( t \) table: area in lower tail is between .05 and .10; therefore, \( p \)-value is between .10 and .20.
Using Excel: \( p \)-value = 2*T.DIST(-1.42,19,TRUE) = .1717
Using unrounded standard deviation and resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .1730

\( p \)-value >.10; Cannot reject \( H_0 \). There is no significant difference between the mean scores for the first and fourth rounds.

b. \( \bar{d} = -1.05 \); First round scores were lower than fourth round scores.

c. \( \alpha = .10 \quad df = 19 \quad t = 1.729 \)

Margin of error = \( t_{0.05} = \frac{s_d}{\sqrt{n}} = 1.729 \frac{3.3162}{\sqrt{20}} = 1.28 \)

Yes, just check to see if the 90\% confidence interval includes a difference of zero. If it does, the difference is not statistically significant.

90\% Confidence interval: -1.05 ± 1.28 (-2.33, .23)

The interval does include 0, so the difference is not statistically significant.
28. a. \( \bar{p}_1 - \bar{p}_2 = .48 - .36 = .12 \)

b. \( \bar{p}_1 - \bar{p}_2 \pm z_{.05} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \)

\[ .12 \pm 1.645 \sqrt{\frac{.48(1-.48)}{400} + \frac{.36(1-.36)}{300}} \]

\[ .12 \pm .0614 \quad (.0586 \text{ to } .1814) \]

c. \( .12 \pm 1.96 \sqrt{\frac{.48(1-.48)}{400} + \frac{.36(1-.36)}{300}} \)

\[ .12 \pm .0731 \quad (.0469 \text{ to } .1931) \]

29. a. \( \bar{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{200(.22) + 300(.16)}{200 + 300} = .1840 \)

\[ z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.22 - .16}{\sqrt{.1840(1-.1840)\left(\frac{1}{200} + \frac{1}{300}\right)}} = 1.70 \]

\( p \)-value is upper-tail area

Using normal table with \( z = 1.70 \): \( p \)-value = 1.0000 - .9554 = .0446
Using Excel: \( p \)-value = 1 - NORM.S.DIST(1.70,TRUE) = .0446
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0449

b. \( p \)-value \( \leq .05 \); reject \( H_0 \).

30. Let \( p_1 \) = the population proportion of executives in Current survey thinking favorable outlook

\( p_2 \) = the population proportion of executives in Prior Year survey thinking favorable outlook

\( \bar{p}_1 = 220/400 = .55 \quad \bar{p}_2 = 192/400 = .48 \)

\( \bar{p}_1 - \bar{p}_2 \pm z_{.025} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \)

\[ .55 - .48 \pm 1.96 \sqrt{\frac{.55(1-.55)}{400} + \frac{.48(1-.48)}{400}} \]

\[ .07 \pm .0691 \quad (.0009 \text{ to } .1391) \]

7\% more executives are predicting an increase in full-time jobs. The confidence interval shows the difference may be from 0\% to 14\%.
32. Let $p_1$ = the population proportion of tuna that is mislabeled
   $p_2$ = the population proportion of mahimahi that is mislabeled

   a. The point estimate of the proportion of tuna that is mislabeled is $\overline{p}_1 = \frac{99}{220} = .45$
   
   b. The point estimate of the proportion of mahimahi that is mislabeled is $\overline{p}_2 = \frac{56}{160} = .35$

   c. $\overline{p}_1 - \overline{p}_2 = .45 - .35 = .10$

   $$\overline{p}_1 - \overline{p}_2 \pm z_{0.025} \sqrt{\frac{\overline{p}_1(1-\overline{p}_1)}{n_1} + \frac{\overline{p}_2(1-\overline{p}_2)}{n_2}}$$

   $$\overline{p}_1 - \overline{p}_2 \pm 1.96 \sqrt{\frac{.45 (.55)}{220} + \frac{.35 (.65)}{160}}$$

   $$\overline{p}_1 - \overline{p}_2 \pm .0989 \quad (.0011 \text{ to } .1989)$$

   The 95% confidence interval estimate of the difference between the proportion of tuna and mahimahi that is mislabeled is $\overline{p}_1 - \overline{p}_2 = .10 \pm .0989$ or (.0011 to .1989).

   With 95% confidence the proportion of mislabeled tuna exceeds the proportion of mislabeled mahimahi by 0% to 20%.

34. Let $p_1$ = the population proportion of wells drilled in 2005 that were dry
   $p_2$ = the population proportion of wells drilled in 2012 that were dry

   a. $H_0: p_1 - p_2 \leq 0$
      $H_a: p_1 - p_2 > 0$ \quad Research hypothesis

   b. $\overline{p}_1 = \frac{24}{119} = .2017$

   c. $\overline{p}_2 = \frac{18}{162} = .1111$

   d. $\overline{p} = \frac{n_1\overline{p}_1 + n_2\overline{p}_2}{n_1 + n_2} = \frac{24 + 18}{119 + 162} = .1495$

   $$z = \frac{\overline{p}_1 - \overline{p}_2}{\sqrt{\overline{p}(1-\overline{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{.2017 - .1111}{\sqrt{.1495(1-.1495) \left( \frac{1}{119} + \frac{1}{162} \right)}} = 2.10$$

   $p$-value is upper-tail area

   Using normal table with $z = 2.10$: \( p \)-value = 1.0000 - .9821 = .0179
   Using Excel: \( p \)-value = 1-NORM.S.DIST(2.10,TRUE) = .0179
   Using unrounded proportions and resulting unrounded Test Statistic via Excel with cell referencing, $p$-value = .0177

   $p$-value < .05, so reject $H_0$ and conclude that wells drilled in 2005 were dry more frequently than wells drilled in 2012. That is, the frequency of dry wells has decreased over the eight years from 2005 to 2012.
36. a. Let \( p_1 \) = population proportion of rooms occupied for current year
\( p_2 \) = population proportion of rooms occupied for previous year

\[ H_0: p_1 - p_2 \leq 0 \]
\[ H_a: p_1 - p_2 > 0 \] Research hypothesis

b. \( \bar{p}_1 = \frac{1470}{1750} = .84 \) (current year)
\( \bar{p}_2 = \frac{1458}{1800} = .81 \) (previous year)

c. \( \bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{1750(.84) + 1800(.81)}{1750 + 1800} = .8248 \)

\[ z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.84 - .81}{\sqrt{.8248(1 - .8248)\left(\frac{1}{1750} + \frac{1}{1800}\right)}} = 2.35 \]

\( p \)-value is upper tail area

Using normal table with \( z = 2.35 \): \( p \)-value = 1.0000 - .9906 = .0094
Using Excel: \( p \)-value = 1 - NORM.S.DIST(2.35,TRUE) = .0094
Using unrounded pooled proportion and resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0094

\( p \)-value \leq .05, reject \( H_0 \). There has been an increase in the hotel occupancy rate.

d. \( \bar{p}_1 - \bar{p}_2 \pm z_{.025} \sqrt{\frac{\bar{p}_1(1 - \bar{p}_1)}{n_1} + \frac{\bar{p}_2(1 - \bar{p}_2)}{n_2}} \)

\[ .84 - .81 \pm 1.96 \sqrt{\frac{.84(1 - .84)}{1750} + \frac{.81(1 - .81)}{1800}} \]

\[ .03 \pm .025 \] (.005 to .055)

Officials would likely be pleased with the occupancy statistics. The trend for the current year is an increase in hotel occupancy rates compared to last year. The point estimate is a 3% increase with a 95% confidence interval from .5% to 5.5%.

38. \( H_0: \mu_1 - \mu_2 = 0 \)
\( H_a: \mu_1 - \mu_2 \neq 0 \) Research hypothesis

\[ z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(4.1 - 3.4) - 0}{\sqrt{\frac{2.2^2}{120} + \frac{1.5^2}{100}}} = 2.79 \]

Because \( z > 0 \), \( p \)-value is two times the upper tail area

Using normal table with \( z = 2.79 \): \( p \)-value = 2(1 - .9974) = .0052
Using Excel: \( p \)-value = 2*(1 - NORM.S.DIST(2.79,TRUE)) = .0052
Using unrounded Test Statistic via Excel with cell referencing, p-value = .0052

\[ p-value \leq .05, \text{ reject } H_0. \text{ A difference exists with system B having the lower mean checkout time.} \]

40. \( \mu_1 \) = population mean return of Load funds
\( \mu_2 \) = population mean return of No Load funds

a. \( H_0 : \mu_1 - \mu_2 \leq 0 \)
\( H_a : \mu_1 - \mu_2 > 0 \) Research hypothesis

b. Using the Mutual data file and Excel’s Data Analysis Tool, the Two Sample t-Test with Unequal Variances results are:

\[
\begin{array}{l|c|c}
& \text{Load Return} & \text{No Load Return} \\
\hline
\text{Mean} & 16.22566667 & 15.70466667 \\
\text{Variance} & 12.38831506 & 10.98724644 \\
\text{Observations} & 30 & 30 \\
\text{Hypothesized Mean Difference} & 0 & \\
\text{df} & 58 & \\
\text{t Stat} & 0.590224625 & \\
\text{P(T<=t) one-tail} & 0.278666388 & \\
\text{t Critical one-tail} & 1.671552762 & \\
\text{P(T<=t) two-tail} & 0.557332776 & \\
\text{t Critical two-tail} & 2.001717484 & \\
\end{array}
\]

\[ n_1 = 30 \quad n_2 = 30 \]
\[ \bar{x}_1 = 16.226 \quad \bar{x}_2 = 15.705 \]
\[ s_1 = 3.52 \quad s_2 = 3.31 \]
\[ t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(16.226 - 15.705) - 0}{\sqrt{\frac{(3.52)^2}{30} + \frac{(3.31)^2}{30}}} = .59 \]
\[
\begin{align*}
\text{df} &= \frac{\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right) + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)} \\
&= \frac{\left(\frac{3.52^2}{30}\right)^2 + \left(\frac{3.31^2}{30}\right)^2}{\frac{1}{29}\left(\frac{3.52^2}{30}\right) + \frac{1}{29}\left(\frac{3.31^2}{30}\right)} = 57.8
\end{align*}
\]

Degrees of freedom = 57
\( p\)-value is upper-tail area

Using \( t \) table: \( p\)-value is greater than .20
Using Excel: \( p\)-value = 1-T.DIST(.59,57,TRUE) = .2788
Using unrounded means and standard deviations and the resulting unrounded Test Statistic via Excel with cell referencing, \( p\)-value = .2787

\[ p\)-value > .05, do not reject \( H_0. \text{ Cannot conclude that the mutual funds with a load have a greater mean rate of return.} \]
42. a. \( x_1 \) = SAT score for twin raised with No Siblings
   \( x_2 \) = SAT score for twin raised with at least one Sibling

Let \( d_i = \text{SAT score for twin raised with No Siblings} - \text{SAT score for twin raised with Siblings} \)

Using the Twins data file and Excel’s Data Analysis Tool, Descriptive Statistics results of the differences are:

<table>
<thead>
<tr>
<th>No Sibling SAT - Siblings SAT Difference</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>

\[
\bar{d} = \frac{\sum d_i}{n} = \frac{280}{20} = 14
\]

b. \( s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{54.880}{19}} = 53.744 \)

\( df = n - 1 = 19, \; t_{0.05} = 1.729 \)

\[
\bar{d} \pm t_{0.05} \frac{s_d}{\sqrt{n}} = 14 \pm 1.729 \frac{53.744}{\sqrt{20}} = 14 \pm 20.78 (-6.78 \text{ to } 34.78)
\]

c. \( H_0: \mu_d = 0 \)
   \( H_a: \mu_d \neq 0 \)  Research hypothesis

Using the Twins data file and Excel’s Data Analysis Tool, Descriptive Statistics and Paired t test results are:

<table>
<thead>
<tr>
<th>t-Test: Paired Two Sample for Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Variance</td>
</tr>
</tbody>
</table>
Chapter 10

### Observations

<table>
<thead>
<tr>
<th>Hypothesized Mean Difference</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>19</td>
</tr>
<tr>
<td>t Stat</td>
<td>1.164964745</td>
</tr>
<tr>
<td>P(T&lt;=t) one-tail</td>
<td>0.129225848</td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>2.539483191</td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.258451697</td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.860934606</td>
</tr>
</tbody>
</table>

\[
t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{14 - 0}{53.744 / \sqrt{20}} = 1.165
\]

Degrees of freedom = \( n - 1 = 19 \)

Because \( t > 0 \), \( p \)-value is two times the upper tail area

Using \( t \) table; area in upper tail is between .10 and .20; therefore, \( p \)-value is between .20 and .40.

Using Excel: \( p \)-value = 2*(1-T.DIST(1.165,19,TRUE)) = .2584

Using unrounded standard deviations and resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .2585

\( p \)-value > .01, do not reject \( H_0 \), cannot conclude that there is a difference between the mean scores for theno sibling and with sibling groups.

44. a. \[ H_0: p_1 - p_2 = 0 \]
   \[ H_a: p_1 - p_2 \neq 0 \]

Research Hypothesis

\[ \bar{p}_1 = 76/400 = .19 \]

\[ \bar{p}_2 = 90/900 = .10 \]

\[ \bar{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{76 + 90}{400 + 900} = .1277 \]

\[ z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = \frac{.19 - .10}{\sqrt{.1277(1-.1277) \left( \frac{1}{400} + \frac{1}{900} \right)}} = 4.49 \]

Because \( z > 0 \), \( p \)-value is two times the upper tail area

Using normal table with \( z = 4.49 \); \( p \)-value \( \approx 2(1 - 1) \approx 0 \)

Using Excel: \( p \)-value \( \approx 2*(1-NORM.S.DIST(4.49,TRUE)) \approx 0 \)

Using unrounded pooled proportion and resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value \( \approx 0 \)

\( p \)-value \( \leq .05 \), Reject \( H_0 \); there is a difference between claim rates.

b. \[ \bar{p}_1 - \bar{p}_2 \pm z_{.05} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \]
   \[ .19 - .10 \pm 1.96 \sqrt{\frac{.19(1-.19)}{400} + \frac{.10(1-.10)}{900}} \]
.09 ± .0432 (.0468 to .1332)

Claim rates are higher for single males.

46. Let \( p_1 \) = the population proportion of American adults under 30 years old
\[ p_2 = \text{the population proportion of Americans who are at least 30 years old} \]

a. From datafileComputerNews, there are 109 Yes responses for each age group. The total number of respondents of the under 30 years group is 200, while the 30 and over group had 150 respondents.

American adults under 30 years old: \( \bar{p}_1 = \frac{109}{200} = .545 \)

Americans who are at least 30 years old: \( \bar{p}_2 = \frac{109}{150} = .727 \)

b. \( \bar{p}_1 - \bar{p}_2 = .545 - .727 = -.182 \)

\[
\sqrt{\frac{.545(1-.545)}{200} + \frac{.727(1-.727)}{150}} = .0506
\]

Confidence interval: -.182 ± 1.96(.0506) or -.182 ± .0992 (-.2809 to -.0824)

c. Since the confidence interval in part (b) does not include 0 and both values are negative, conclude that the proportion of American adults under 30 years old who use a computer to gain access to news is less than the proportion of Americans who are at least 30 years old that use a computer to gain access to news.
Chapter 11
Inferences About Population Variances

Solutions:

2. \( s^2 = 25, \ n = 20, \ df = 19 \)

    a. For 90% confidence, \( \chi^2_{0.05} = 30.144 \) and \( \chi^2_{0.95} = 10.117 \)

\[
\frac{(n-1)s^2}{\chi^2_{a/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-a/2}}
\]

\[
\frac{19(25)}{30.144} \leq \sigma^2 \leq \frac{19(25)}{10.117}
\]

\[
15.76 \leq \sigma^2 \leq 46.95
\]

b. For 95% confidence, \( \chi^2_{0.025} = 32.852 \) and \( \chi^2_{0.975} = 8.907 \)

\[
\frac{(n-1)s^2}{\chi^2_{a/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-a/2}}
\]

\[
\frac{19(25)}{32.852} \leq \sigma^2 \leq \frac{19(25)}{8.907}
\]

\[
14.46 \leq \sigma^2 \leq 53.33
\]

c. \( \sqrt{\frac{(n-1)s^2}{\chi^2_{a/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-a/2}}} \)

\[
\sqrt{14.46} \leq \sigma \leq \sqrt{53.33}
\]

\[
3.8 \leq \sigma \leq 7.3
\]

4. \( s^2 = .36, \ n = 18, \ df = 17 \)

    a. For 90% confidence, \( \chi^2_{0.05} = 27.587 \) and \( \chi^2_{0.95} = 8.672 \)

\[
\frac{(n-1)s^2}{\chi^2_{a/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-a/2}}
\]

\[
\frac{17(.36)}{27.587} \leq \sigma^2 \leq \frac{17(.36)}{8.672}
\]
Chapter 11

\[ .22 \leq \sigma^2 \leq .71 \]

b. \[ \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}} \]

\[ \sqrt{.22} \leq \sigma \leq \sqrt{.71} \]

\[ .47 \leq \sigma \leq .84 \]

6. a. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{656}{16} = 41 \]

The sample mean amount spent on a Halloween costume was $41.

b. \[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{8298}{16-1} = 553.2 \]

\[ s = \sqrt{553.2} = 23.52 \]

c. \[ s^2 = 553.2, \ n = 16, \ \text{df} = 15 \]

For 95% confidence, \( \chi^2_{.025} = 27.488 \) and \( \chi^2_{.975} = 6.262 \)

\[ \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \]

\[ \frac{16-1)(553.2)}{27.488} \leq \sigma^2 \leq \frac{16-1)(553.2)}{6.262} \]

\[ 301.88 \leq \sigma^2 \leq 1325.14 \]

Note: if using excel functions to determine chi square critical values and cell referencing for unrounded answers, the interval is \[ 301.88 \leq \sigma^2 \leq 1325.11 \]

8. a. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{9.36}{12} = .78 \]

\[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{5.2230}{12-1} = .4748 \]

b. \[ s = \sqrt{.4748} = .6891 \]

c. \[ s^2 = .4748, \ n = 12, \ \text{df} = 11 \]
Inferences About Population Variances

For 95% confidence, $\chi^2_{0.025} = 21.920$ and $\chi^2_{0.075} = 3.816$

\[
\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}
\]

\[
\frac{(12-1).4748}{21.920} \leq \sigma^2 \leq \frac{(12-1).4748}{3.816}
\]

$.2383 \leq \sigma^2 \leq 1.3687$

Note: if using excel functions to determine chi square critical values and cell referencing for unrounded answers, the interval is $.2383 \leq \sigma^2 \leq 1.3688$

\[
\frac{\sqrt{(n-1)s^2}}{\chi^2_{\alpha/2}} \leq \sigma \leq \frac{\sqrt{(n-1)s^2}}{\chi^2_{1-\alpha/2}}
\]

\[
\sqrt{.2383} \leq \sigma \leq \sqrt{1.3687}
\]

$.4882 \leq \sigma \leq 1.1699$

Note: if using excel functions to determine chi square critical values and cell referencing for unrounded answers, the interval is $.4881 \leq \sigma \leq 1.1700$

9. $H_0: \sigma^2 \leq .0004$

$H_a: \sigma^2 > .0004$

Research hypothesis

\[
\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1)(.0005)}{.0004} = 36.25
\]

Degrees of freedom = $n - 1 = 29$

$p$-value is upper-tail area

Using $\chi^2$ table, $p$-value is greater than .10

Using Excel, $p$-value = CHISQ.DIST.RT(36.25,29) = .1664

$p$-value > .05, do not reject $H_0$. The product specification does not appear to be violated.

10. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{1260}{15} = 84$

b. $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{1662}{15 - 1} = 118.71$

c. $s = \sqrt{s^2} = \sqrt{118.71} = 10.90$

d. Hypothesis for $\sigma = 12$ is for $\sigma^2 = (12)^2 = 144$

$H_0: \sigma^2 = 144$

$H_a: \sigma^2 \neq 144$
Degrees of freedom = \( n-1 = 14 \)

Because the left tail is the nearest tail in this 2 tail test, the \( p \)-value is 2 * the lower tail area

Using \( \chi^2 \) table, area in the lower tail is greater than (1 – .90) = .10; therefore, \( p \)-value is greater than .20

Using Excel, \( p \)-value corresponding to \( \chi^2 = 11.54 \) is = 2*CHISQ.DIST(11.54,14,TRUE) 
= 2*.3568 = .7136

Using unrounded standard deviation and the resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .7139

\( p \)-value > .05, do not reject \( H_0 \). The hypothesis that the population standard deviation is 12 cannot be rejected

12. a. \( s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = \frac{8.9167}{11} = .8106 \)

b. \( H_0: \sigma^2 = .94 \)
\( H_a: \sigma^2 \neq .94 \)

\( \chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(12-1)(.8106)}{.94} = 9.49 \)

Degrees of freedom = \( n - 1 = 11 \)

Because the left tail is the nearest tail in this 2 tail test, the \( p \)-value is 2 * the lower tail area

Using \( \chi^2 \) table, area in the lower tail is greater than (1 – .90) = .10; therefore, \( p \)-value is greater than .20

Using Excel, \( p \)-value = 2*CHISQ.DIST(9.49,11,TRUE) = .8465

Using unrounded standard deviation and the resulting unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .8457

\( p \)-value > .05, cannot reject \( H_0 \).

14. a. Population 1 is the one with the larger variance

\( n_1=16, n_2=21 \)

\( F = \frac{s_1^2}{s_2^2} = \frac{5.8}{2.4} = 2.4 \)

Upper Tail test with Degrees of freedom 15 and 20

Using \( F \) table, \( p \)-value is between .025 and .05

Using Excel, \( p \)-value = F.DIST.RT(2.4,15,20) = .0345

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0334

\( p \)-value \( \leq \) .05, reject \( H_0 \). Conclude \( \sigma_1^2 > \sigma_2^2 \)
b. Critical $F_{0.05} = 2.20$

Reject $H_0$ if $F \geq 2.20$

$2.4 \geq 2.20$, reject $H_0$. Conclude $\sigma_1^2 > \sigma_2^2$

15. a. Larger sample variance is $s_1^2$

$n_1=21, n_2=26$

$$F = \frac{s_1^2}{s_2^2} = \frac{8.2}{4} = 2.05$$

Two Tail test ($2 \times$ right tail) with Degrees of freedom 20 and 25

Using $F$ table, area in the upper tail is between .025 and .05; Two-tail $p$-value is between .05 and .10

Using Excel, $p$-value = $2 \times F.DIST.RT(2.05,20,25) = .0904$

$p$-value > .05, do not reject $H_0$.

b. Since we have a two-tailed test

$F_{0.025} = F_{0.05} = 2.30$

Reject $H_0$ if $F \geq 2.30$

$2.05 < 2.30$, do not reject $H_0$.

16. For this type of hypothesis test, we place the larger variance in the numerator. So the Fidelity variance is given the subscript of 1. $s_1 = 18.9, s_2 = 15$

$H_0: \sigma_1^2 \leq \sigma_2^2$

$H_a: \sigma_1^2 > \sigma_2^2$  \hspace{1cm} Research hypothesis

$$F = \frac{s_1^2}{s_2^2} = \frac{18.9^2}{15.0^2} = 1.5876$$

Upper Tail test; Degrees of freedom in the numerator and denominator are both 59

Using the $F$ table and estimating with 60 df for each, $p$-value is between .025 and .05

Using Excel, $p$-value corresponding to $F = 1.5876$ is $F.DIST.RT(1.5876,59,59) = .0392$

$p$-value $\leq .05$, reject $H_0$. We conclude that the Fidelity fund has a greater variance than the American Century fund and therefore is more risky.

17. a. Population 1 is 4 year old automobiles since it is the one with the larger variance

$s_1 = 170, n_1=26, s_2 = 100, n_2=25$

$H_0: \sigma_1^2 \leq \sigma_2^2$

$H_a: \sigma_1^2 > \sigma_2^2$  \hspace{1cm} Research hypothesis
Chapter 11

b. \( F = \frac{s_1^2}{s_2^2} = \frac{170^2}{100^2} = 2.89 \)

Upper Tail test; Degrees of freedom 25 and 24

Using \( F \) table, \( p \)-value is less than .01
Using Excel, \( p \)-value = F.DIST.RT(2.89,25,24) = .0057

\( p \)-value \( \leq .01 \), reject \( H_0 \). Conclude that 4 year old automobiles have a larger variance in annual repair costs compared to 2 year old automobiles. This is expected due to the fact that older automobiles are more likely to have some more expensive repairs which lead to greater variance in the annual repair costs.

18. We place the larger sample variance in the numerator. So, the Merrill Lynch variance is given the subscript of 1. \( s_1^2 = 587, \ n_1=16, \ s_2 = 489, \ n_2=10 \)

\( H_0: \sigma_1^2 = \sigma_2^2 \)
\( H_a: \sigma_1^2 \neq \sigma_2^2 \)  Research hypothesis

\( F = \frac{s_1^2}{s_2^2} = \frac{587^2}{489^2} = 1.44 \)

Two Tail test (2 * right tail) with Degrees of freedom 15 and 9

Using \( F \) table, area in the upper tail is greater than .10; Two-tail \( p \)-value is greater than .20

Using Excel, \( p \)-value corresponding to \( F = 1.44 \) is 2*F.DIST.RT(1.44,15,9) = .5906
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .5898

\( p \)-value > .10, do not reject \( H_0 \). We cannot conclude there is a statistically significant difference between the variances for the two companies.

20. Population 1 is Managers since it is the one with the larger variance \( s_1^2 = 11.1, \ n_1=26, \ s_2 = 2.1, \ n_2=25 \)

\( H_0: \sigma_1^2 = \sigma_2^2 \)
\( H_a: \sigma_1^2 \neq \sigma_2^2 \)

\( F = \frac{s_1^2}{s_2^2} = \frac{11.1}{2.1} = 5.29 \)

Two Tail test (2 * right tail) with Degrees of freedom 25 and 24

Using \( F \) table, area in the upper tail is less than .01; Two-tail \( p \)-value is less than .02

Using Excel, \( p \)-value = 2*F.DIST.RT(5.29,25,24) = .0001

\( p \)-value \( \leq .05 \), reject \( H_0 \). The population variances are not equal for seniors and managers.
22. a. Since it is the one with the larger variance, Population 1 - Wet pavement.
   $s_1 = 32, \ n_1=16, \ s_2 = 16, \ n_2=16$

   $H_0: \sigma_1^2 \leq \sigma_2^2$
   $H_a: \sigma_1^2 > \sigma_2^2$  \hspace{1cm} Research hypothesis

   \[ F = \frac{s_1^2}{s_2^2} = \frac{32^2}{16^2} = 4.00 \]

   Upper Tail test; Degrees of freedom 15 and 15

   Using $F$ table, $p$-value is less than .01

   Using Excel, $p$-value = F.DIST.RT(4.00,15,15) = .0054

   $p$-value $\leq .05$, reject $H_0$. Conclude that there is greater variability in stopping distances on wet pavement.

b. Drive carefully on wet pavement because of the uncertainty in stopping distances.

24. $s = 14.95, \ n=13, \ df = 12$

   For 95% confidence, $\chi^2_{0.025} = 23.337$ and $\chi^2_{0.075} = 4.404$

   \[ \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \]

   \[ \frac{(12)(14.95)^2}{23.337} \leq \sigma^2 \leq \frac{(12)(14.95)^2}{4.404} \]

   $114.9 \leq \sigma^2 \leq 609$

   \[ \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}} \]

   $\sqrt{114.9} \leq \sigma \leq \sqrt{609}$

   $10.72 \leq \sigma \leq 24.68$

26. a. $s = .014, \ n=15, \ df = 14$

   $H_0: \sigma^2 \leq .0001$
   $H_a: \sigma^2 > .0001$  \hspace{1cm} Research hypothesis

   \[ \chi^2 = \frac{(n-1)s^2}{\sigma_a^2} = \frac{(15-1)(0.014)^2}{.0001} = 27.44 \]

   Degrees of freedom $= n - 1 = 14$

   $p$-value is upper-tail area

   Using $\chi^2$ table, $p$-value is between .01 and .025
Using Excel, \( p\)-value = CHISQ.DIST.RT(27.44,14) = .0169

\( p\)-value \leq .10, reject \( H_0 \). Variance exceeds maximum variance requirement.

b. For 90% confidence, \( df = 14 \), \( \chi^2_{.05} = 23.685 \) and \( \chi^2_{.95} = 6.571 \)

\[
\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}
\]

\[
\frac{(14)(.014)^2}{23.685} \leq \sigma^2 \leq \frac{(14)(.014)^2}{6.571}
\]

\[.00012 \leq \sigma^2 \leq .00042\]

28. \( s^2 = 1.5, n = 22, \ df = 21 \)

\( H_0: \sigma^2 \leq 1 \)
\( H_a: \sigma^2 > 1 \)

\[
\chi^2 = \frac{(n-1)s^2}{\sigma^2_0} = \frac{(22-1)(1.5)}{1} = 31.50
\]

Degrees of freedom = \( n - 1 = 21 \)

\( p\)-value is upper-tail area

Using \( \chi^2 \) table, \( p\)-value is between .05 and .10

Using Excel, \( p\)-value = CHISQ.DIST.RT(31.50,21) = .0657

\( p\)-value \leq .10, reject \( H_0 \). Conclude that \( \sigma^2 > 1 \).

30. a. Try \( n = 15 \) with \( s = 8 \),

For 95% confidence, \( df = 14 \), \( \chi^2_{.025} = 26.119 \) and \( \chi^2_{.975} = 5.629 \)

\[
\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}
\]

\[
\frac{(14)(64)}{26.119} \leq \sigma^2 \leq \frac{(14)(64)}{5.629}
\]

\[34.3 \leq \sigma^2 \leq 159.2\]

\[
\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}
\]

\[\sqrt{34.3} \leq \sigma \leq \sqrt{159.2}\]

\[5.86 \leq \sigma \leq 12.62\]

Therefore, a sample size of 15 was used.
b. \(n = 25\); expect the width of the interval to be smaller.

For 95\% confidence, \(df = 24\), \(\chi^2_{0.025} = 39.364\) and \(\chi^2_{0.975} = 12.401\)

\[
\frac{(n-1)s^2}{\chi^2_{0.025}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.975}}
\]

\[
\frac{(24)(8)^2}{39.364} \leq \sigma^2 \leq \frac{(24)(8)^2}{12.401}
\]

\[
39.02 \leq \sigma^2 \leq 123.86
\]

\[
\sqrt{39.02} \leq \sigma \leq \sqrt{123.86}
\]

\[
6.25 \leq \sigma \leq 11.13
\]

32. \(H_0: \sigma_1^2 = \sigma_2^2\)

\(H_a: \sigma_1^2 \neq \sigma_2^2\)  

Research hypothesis

Since it is the one with the larger variance, population 1 is those who completed course 
\(s_1 = .940, \ n_1=352, \ s_2 = .797, \ n_2=73\)

Using critical value approach and Excel to determine the critical value, since \(F\) tables do not have 351 and 72 degrees of freedom.

Using Excel, \(F.INV.RT(0.025,351,72)\) gives the Critical \(F_{0.025} = 1.466\)

Reject \(H_0\) if \(F \geq 1.466\)

\[
F = \frac{s_1^2}{s_2^2} = \frac{.940^2}{.797^2} = 1.39
\]

\(F < 1.466\), do not reject \(H_0\). We are not able to conclude students who complete the course and students who drop out have different variances of grade point averages.

Using Excel, the p-value approach gives the following:

Two Tail test (2 * right tail) with Degrees of freedom 351 and 72

Using Excel: \(p\)-value = 2*F.DIST.RT(1.39,351,72) = .0906

Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .0899

\(p\)-value > .05, do not reject \(H_0\). There is not a statistically significant difference in the variances.
34. $H_0: \sigma_1^2 = \sigma_2^2$
$H_a: \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2} = \frac{25}{12} = 2.08$$

Two Tail test (2 * right tail) with Degrees of freedom 30 and 24

Using $F$ table, area in tail is between .025 and .05; Two-tail $p$-value is between .05 and .10

Using Excel, $p$-value = 2*F.DIST.RT(2.08,30,24) = .0695
Using unrounded Test Statistic via Excel with cell referencing, $p$-value = .0689

$p$-value $\leq .10$, reject $H_0$. Conclude that the population variances are not equal.
Chapter 12
Tests of Goodness of Fit and Independence

Solutions:

1. a. Expected frequencies: \( e_1 = 200 \cdot (0.40) = 80 \), \( e_2 = 200 \cdot (0.40) = 80 \)
   
   \( e_3 = 200 \cdot (0.20) = 40 \)

   Actual frequencies: \( f_1 = 60, f_2 = 120, f_3 = 20 \)

   \[
   \chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}
   \]

   \[
   \chi^2 = \frac{(60 - 80)^2}{80} + \frac{(120 - 80)^2}{80} + \frac{(20 - 40)^2}{40}
   \]

   \[
   = \frac{400}{80} + \frac{1600}{80} + \frac{400}{40}
   \]

   \[
   = 5 + 20 + 10
   \]

   \[
   = 35
   \]

   \( k - 1 = 2 \) degrees of freedom

   Using the \( \chi^2 \) table with \( df = 2 \), \( \chi^2 = 35 \) shows the \( p \)-value is less than .005.

   Using Excel, the \( p \)-value corresponding to \( \chi^2 = 35 \): \( = \text{CHISQ.DIST.RT}(35,2) = .0000 \)

   \( p \)-value \( \leq .01 \), reject \( H_0 \)

   b. For 2 degrees of freedom, \( \chi_{0.01} = 9.210 \)

   Reject \( H_0 \) if \( \chi^2 \geq 9.210 \)

   \( \chi^2 = 35 \), reject \( H_0 \)

2. Expected frequencies: \( e_1 = 300 \cdot (0.25) = 75 \), \( e_2 = 300 \cdot (0.25) = 75 \)
   
   \( e_3 = 300 \cdot (0.25) = 75 \), \( e_4 = 300 \cdot (0.25) = 75 \)

   Actual frequencies: \( f_1 = 85, f_2 = 95, f_3 = 50, f_4 = 70 \)

   \[
   \chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}
   \]
Using the $\chi^2$ table with $df = 3, \chi^2 = 15.33$ shows the $p$-value is less than .005.
Using Excel, the $p$-value corresponding to $\chi^2 = 15.33; = \text{CHISQ.DIST.RT}(15.33,3) = .0016$.

$p$-value $\leq .05$, reject $H_0$.

The population proportions are not the same.

3. $H_0 = p_{ABC} = .29, p_{CBS} = .28, p_{NBC} = .25, p_{IND} = .18$

$H_a = $ The proportions are not $p_{ABC} = .29, p_{CBS} = .28, p_{NBC} = .25, p_{IND} = .18$

Expected frequencies: 300 (.29) = 87, 300 (.28) = 84
300 (.25) = 75, 300 (.18) = 54
$e_1 = 87, e_2 = 84, e_3 = 75, e_4 = 54$

Actual frequencies: $f_1 = 95, f_2 = 70, f_3 = 89, f_4 = 46$

$$\chi^2 = \sum (f_i - e_i)^2 / e_i$$

$$\chi^2 = \frac{(95 - 87)^2}{87} + \frac{(70 - 84)^2}{84} + \frac{(89 - 75)^2}{75} + \frac{(46 - 54)^2}{54}$$

$$= 6.87$$

$k - 1 = 3$ degrees of freedom

Using the $\chi^2$ table with $df = 3, \chi^2 = 6.87$ shows the $p$-value is between .05 and .10.
Using Excel, the $p$-value corresponding to $\chi^2 = 6.87; = \text{CHISQ.DIST.RT}(6.87,3) = .0762$.

$p$-value $> .05$, do not reject $H_0$. There has not been a significant change in the viewing audience proportions.
4. \( H_0: \) Color proportions are .24 Blue, .13 Brown, .2 Green, .16 Orange, .13 Red and .14 Yellow  
\( H_a: \) Color proportions differ from the above

\[
\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}
\]

<table>
<thead>
<tr>
<th>Category</th>
<th>Hypothesized Proportion (p)</th>
<th>Observed Frequency ((f_i))</th>
<th>Expected Frequency ((e_i) = n^*p)</th>
<th>Chi Square ((f_i - e_i)^2 / e_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>.24</td>
<td>105</td>
<td>120</td>
<td>1.88</td>
</tr>
<tr>
<td>Brown</td>
<td>.13</td>
<td>72</td>
<td>65</td>
<td>.75</td>
</tr>
<tr>
<td>Green</td>
<td>.20</td>
<td>89</td>
<td>100</td>
<td>1.21</td>
</tr>
<tr>
<td>Orange</td>
<td>.16</td>
<td>84</td>
<td>80</td>
<td>.20</td>
</tr>
<tr>
<td>Red</td>
<td>.13</td>
<td>70</td>
<td>65</td>
<td>.38</td>
</tr>
<tr>
<td>Yellow</td>
<td>.14</td>
<td>80</td>
<td>70</td>
<td>1.43</td>
</tr>
<tr>
<td>Total:</td>
<td>500 = n</td>
<td></td>
<td></td>
<td>( \chi^2 = 5.85 )</td>
</tr>
</tbody>
</table>

\( k - 1 = 6 - 1 = 5 \) degrees of freedom  
Using the \( \chi^2 \) table with \( df = 5 \), \( \chi^2 = 5.85 \) shows the \( p \)-value is greater than .10

Using Excel, the \( p \)-value corresponding to \( \chi^2 = 5.85 \): \( \text{CHISQ.DIST.RT(5.85,5)} = .3211 \)

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .3209

\( p \)-value > .05, do not reject \( H_0 \). We cannot reject the hypothesis that the overall percentages of colors in the population of M&M milk chocolate candies are .24 blue, .13 brown, .20 green, .16 orange, .13 red and .14 yellow.

6. a. \( H_0: \) \( p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = \frac{1}{7} \)

\( H_a: \) Not all proportions are equal

<table>
<thead>
<tr>
<th>Observed Frequency ((f_i))</th>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>50</td>
<td>53</td>
<td>47</td>
<td>55</td>
<td>69</td>
<td>80</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expected Frequency ((e_i))</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>60</td>
</tr>
<tr>
<td>Monday</td>
<td>60</td>
</tr>
<tr>
<td>Tuesday</td>
<td>60</td>
</tr>
<tr>
<td>Wednesday</td>
<td>60</td>
</tr>
<tr>
<td>Thursday</td>
<td>60</td>
</tr>
<tr>
<td>Friday</td>
<td>60</td>
</tr>
<tr>
<td>Saturday</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chi Square Calculations ((f_i - e_i)^2 / e_i)</th>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 = \sum \frac{(f_i - e_i)^2}{e_i} )</td>
<td>.60</td>
<td>1.67</td>
<td>.82</td>
<td>2.82</td>
<td>.42</td>
<td>1.35</td>
<td>6.67</td>
</tr>
</tbody>
</table>

\( \chi^2 = \sum \frac{(f_i - e_i)^2}{e_i} = 14.33 \)

\[
\chi^2 = \frac{(66 - 60)^2}{60} + \frac{(50 - 60)^2}{60} + \frac{(53 - 60)^2}{60} + \frac{(47 - 60)^2}{60} + \frac{(55 - 60)^2}{60} + \frac{(69 - 60)^2}{60} + \frac{(80 - 60)^2}{60} = 14.33
\]

Degrees of freedom = \((k - 1) = (7 - 1) = 6\)

Using the \( \chi^2 \) table with \( df = 6 \), \( \chi^2 = 14.33 \) shows the \( p \)-value is between .025 and .05.
Using Excel, the \( p \)-value corresponding to \( \chi^2 = 14.33;= \text{CHISQ.DIST.RT}(14.33,6) = .0262 \)

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0261

\( p \)-value \( \leq .05 \), reject \( H_0 \). Conclude the proportion of traffic accidents is not the same for each day of the week.

b. Percentage of traffic accidents by day of the week

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>66/420  = .1571</td>
<td>15.71%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monday</td>
<td>50/420 = .1190</td>
<td>11.90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday</td>
<td>53/420 = .1262</td>
<td>12.62%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday</td>
<td>47/420 = .1119</td>
<td>11.19%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday</td>
<td>55/420 = .1310</td>
<td>13.10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friday</td>
<td>69/420 = .1643</td>
<td>16.43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturday</td>
<td>80/420 = .1905</td>
<td>19.05%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Saturday has the highest percentage of traffic accident (19%). Saturday is typically the late night and more social day/evening of the week. Alcohol, speeding and distractions are more likely to affect driving on Saturdays. Friday is the second highest with 16.43%.

7. \( H_0 \): The column variable is independent of the row variable

\( H_a \): The column variable is not independent of the row variable

Observed Frequencies \((f_{ij})\)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>20</td>
<td>44</td>
<td>50</td>
<td>114</td>
</tr>
<tr>
<td>Q</td>
<td>30</td>
<td>26</td>
<td>30</td>
<td>86</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>70</td>
<td>80</td>
<td>200</td>
</tr>
</tbody>
</table>

Expected Frequencies \((e_{ij})\)

\[ e_{ij} = \frac{(Row_i \times Total)(Column_j \times Total)}{SampleSize} \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>28.5</td>
<td>39.9</td>
<td>45.6</td>
<td>114</td>
</tr>
<tr>
<td>Q</td>
<td>21.5</td>
<td>30.1</td>
<td>34.4</td>
<td>86</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>70</td>
<td>80</td>
<td>200</td>
</tr>
</tbody>
</table>

Chi-Square Calculations \((f_{ij} - e_{ij})^2 / e_{ij}\)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>2.54</td>
<td>.42</td>
<td>.42</td>
<td>3.38</td>
</tr>
<tr>
<td>Q</td>
<td>3.36</td>
<td>.56</td>
<td>.56</td>
<td>4.48</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 7.86 \]

Degrees of freedom = \((r - 1)(c - 1) = (2-1)(3-1) = 2\)

Using the \( \chi^2 \) table with \( df = 2 \), \( \chi^2 = 7.86 \) shows the \( p \)-value is between .01 and .025.

Using Excel, the \( p \)-value corresponding to \( \chi^2 = 7.86;= \text{CHISQ.DIST.RT}(7.86,2) = .0196 \).

\( p \)-value \( \leq .05 \), reject \( H_0 \). Conclude that there is an association between the column variable and the row variable. The variables are not independent.

8. \( H_0 \): The column variable is independent of the row variable
Tests of Goodness of Fit and Independence

**$H_0$:** The column variable is not independent on the row variable

**Observed Frequencies ($f_{ij}$)**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>20</td>
<td>30</td>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>Q</td>
<td>30</td>
<td>60</td>
<td>25</td>
<td>115</td>
</tr>
<tr>
<td>R</td>
<td>10</td>
<td>15</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>105</td>
<td>75</td>
<td>240</td>
</tr>
</tbody>
</table>

**Expected Frequencies ($e_{ij}$)**

$$e_{ij} = \frac{(Row\ Total)(Column\ Total)}{Sample\ Size}$$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>17.50</td>
<td>30.63</td>
<td>21.88</td>
<td>70</td>
</tr>
<tr>
<td>Q</td>
<td>28.75</td>
<td>50.31</td>
<td>35.94</td>
<td>115</td>
</tr>
<tr>
<td>R</td>
<td>13.75</td>
<td>24.06</td>
<td>17.19</td>
<td>55</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>105</td>
<td>75</td>
<td>240</td>
</tr>
</tbody>
</table>

**Chi–Square Calculations ($f_{ij} - e_{ij}$)$^2 / e_{ij}$**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>.36</td>
<td>.01</td>
<td>.16</td>
<td>.53</td>
</tr>
<tr>
<td>Q</td>
<td>.05</td>
<td>1.87</td>
<td>3.33</td>
<td>5.25</td>
</tr>
<tr>
<td>R</td>
<td>1.02</td>
<td>3.41</td>
<td>9.55</td>
<td>13.99</td>
</tr>
</tbody>
</table>

$$\chi^2 = 19.77 \quad \chi^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Degrees of freedom = $(r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$

Using the $\chi^2$ table with $df = 4$, $\chi^2 = 19.77$ shows the $p$–value is less than .005.

Using Excel, the $p$–value corresponding to $\chi^2 = 19.77$ is $\text{CHISQ.DIST.RT}(19.77,4) = .0006$.

$p$–value $\leq .05$, reject $H_0$. Conclude that the column variable is not independent of the row variable.

**9. a. $H_0$: Type of ticket purchased is independent of the type of flight**

$H_a$: Type of ticket purchased is not independent of the type of flight

**Expected Frequencies: $e_{ij} = \frac{(Row\ Total)(Column\ Total)}{Sample\ Size}$**

<table>
<thead>
<tr>
<th></th>
<th>Ticket</th>
<th>Flight</th>
<th>Observed Frequency ($f_i$)</th>
<th>Expected Frequency ($e_i$)</th>
<th>Chi–square $(f_i - e_i)^2 / e_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>First</td>
<td>Domestic</td>
<td>29</td>
<td>35.59</td>
<td>1.22</td>
</tr>
<tr>
<td>Q</td>
<td>First</td>
<td>International</td>
<td>22</td>
<td>15.41</td>
<td>2.82</td>
</tr>
<tr>
<td>R</td>
<td>Business</td>
<td>Domestic</td>
<td>95</td>
<td>150.73</td>
<td>20.61</td>
</tr>
</tbody>
</table>
Degrees of freedom = \((r - 1)(c - 1) = (3 - 1)(2 - 1) = 2\)

Using the \(\chi^2\) table with \(df = 2\), \(\chi^2 = 100.43\) shows the \(p\)-value is less than .005.

Using Excel, the \(p\)-value corresponding to \(\chi^2 = 100.43\); \(= CHISQ.DIST.RT(100.43,2) = .0000\).

\(p\)-value \(\leq .05\), reject \(H_0\). Conclude that the type of ticket purchased is not independent of the type of flight. We can expect the type of ticket purchased to depend upon whether the flight is domestic or international.

b. Column Percentages = \(\text{Ticket & Flight Type Frequency / Column Total}\); For example: \(29/642 = .045 = 4.5\%

\begin{center}
\begin{tabular}{l|c|c}
Type of Flight & Domestic & International \\
\hline
First Class & 4.5\% & 7.9\% \\
Business Class & 14.8\% & 43.5\% \\
Economy Class & 80.7\% & 48.6\% \\
\hline
\end{tabular}
\end{center}

A higher percentage of first class and business class tickets are purchased for international flights compared to domestic flights. Economy class tickets are purchased more for domestic flights. The first class or business class tickets are purchased for more than 50% of the international flights; 7.9\% + 43.5\% = 51.4\%.

10. a.

\(H_0: \) Employment plan is independent of the type of company

\(H_a: \) Employment plan is not independent of the type of company

\begin{center}
\begin{tabular}{l|c|c|c}
Employment Plan & Private & Public & Total \\
\hline
Add Employees & 37 & 32 & 69 \\
No Change & 19 & 34 & 53 \\
Lay Off Employees & 16 & 42 & 58 \\
Total & 72 & 108 & 180 \\
\hline
\end{tabular}
\end{center}

Expected Frequency \((e_{ij})\)

\[ e_{ij} = \frac{(Row Total)(Column Total)}{Sample Size} \]

\begin{center}
\begin{tabular}{l|c|c|c}
Employment Plan & Private & Public & Total \\
\hline
Add Employees & 27.6 & 41.4 & 69 \\
No Change & 21.2 & 31.8 & 53 \\
Lay Off Employees & 23.2 & 34.8 & 58 \\
Total & 72.0 & 108.0 & 180 \\
\hline
\end{tabular}
\end{center}

Chi Square Calculations \((f_{ij} - e_{ij})^2 / e_{ij}\)

\begin{center}
\begin{tabular}{l|c|c|c}
Employment Plan & Private & Public & Total \\
\hline
\end{tabular}
\end{center}
Tests of Goodness of Fit and Independence

Add Employees 3.20 2.13 5.34
No Change 0.23 0.15 0.38
Lay Off Employees 2.23 1.49 3.72

χ² = 9.44

Degrees of freedom = (r – 1)(c – 1) = (3 – 1)(2 – 1) = 2

Using the χ² table with df = 2, χ² = 9.44 shows the p-value is between .005 and .01.

Using Excel, the p-value corresponding to χ² = 9.44; CHISQ.DIST.RT(9.44,2) = .0089

p-value ≤ .05, reject H₀. Conclude the employment plan is not independent of the type of company. Thus, we expect employment plan to differ for private and public companies.

b. Column probabilities = Company & Plan Frequency/Column Total: For example, 37/72 = .5139

<table>
<thead>
<tr>
<th>Employment Plan</th>
<th>Private</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Employees</td>
<td>.5139</td>
<td>.2963</td>
</tr>
<tr>
<td>No Change</td>
<td>.2639</td>
<td>.3148</td>
</tr>
<tr>
<td>Lay Off Employees</td>
<td>.2222</td>
<td>.3889</td>
</tr>
</tbody>
</table>

Employment opportunities look to be much better for private companies with over 50% of private companies planning to add employees (51.39%). Public companies have the greater proportions of no change and lay off employees planned. 38.89% of public companies are planning to lay off employees over the next 12 months. 69/180 = .3833, or 38.33% of the companies in the survey are planning to hire and add employees during the next 12 months.

12. a. H₀: Quality rating is independent of the education of the owner  
H₁: Quality rating is not independent of the education of the owner

Observed Frequencies (fᵢⱼ)

<table>
<thead>
<tr>
<th>Quality Rating</th>
<th>Some HS</th>
<th>HS Grad</th>
<th>Some College</th>
<th>College Grad</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>35</td>
<td>30</td>
<td>20</td>
<td>60</td>
<td>145</td>
</tr>
<tr>
<td>Outstanding</td>
<td>45</td>
<td>45</td>
<td>50</td>
<td>90</td>
<td>230</td>
</tr>
<tr>
<td>Exceptional</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>50</td>
<td>125</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>500</td>
</tr>
</tbody>
</table>

Expected Frequencies (eᵢⱼ)

\[ e_{ij} = \frac{(Row_i Total)(Column_j Total)}{SampleSize} \]

<table>
<thead>
<tr>
<th>Quality Rating</th>
<th>Some HS</th>
<th>HS Grad</th>
<th>Some College</th>
<th>College Grad</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>58</td>
<td>145</td>
</tr>
<tr>
<td>Outstanding</td>
<td>46</td>
<td>46</td>
<td>46</td>
<td>92</td>
<td>230</td>
</tr>
<tr>
<td>Exceptional</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>50</td>
<td>125</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>500</td>
</tr>
</tbody>
</table>

Chi Square Calculations (fᵢⱼ – eᵢⱼ)² / eᵢⱼ
Chapter 12

### Quality Rating

<table>
<thead>
<tr>
<th>Quality Rating</th>
<th>Some HS</th>
<th>HS Grad</th>
<th>Some College</th>
<th>College Grad</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>1.24</td>
<td>.03</td>
<td>2.79</td>
<td>.07</td>
<td>4.14</td>
</tr>
<tr>
<td>Outstanding</td>
<td>.02</td>
<td>.02</td>
<td>.35</td>
<td>.04</td>
<td>.43</td>
</tr>
<tr>
<td>Exceptional</td>
<td>1.00</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 6.57 \]

Degrees of freedom \((r - 1)(c - 1) = (3 - 1)(4 - 1) = 6\)

Using the \(\chi^2\) table with \(df = 6\), \(\chi^2 = 6.57\) shows the \(p\)-value is greater than .10

Using Excel, the \(p\)-value corresponding to \(\chi^2 = 6.57\): \(= \text{CHISQ.DIST.RT}(6.57,6) = .3624\)

Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .3622

\(p\)-value > .05, do not reject \(H_0\). We are unable to conclude that the quality rating is not independent of the education of the owner. Thus, quality ratings are not expected to differ with the education of the owner.

b. Average: \(145/500 = 29\%\)

Outstanding: \(230/500 = 46\%\)

Exceptional: \(125/500 = 25\%\)

New owners look to be pretty satisfied with their new automobiles with almost 50% rating the quality outstanding and over 70% rating the quality outstanding or exceptional.

14. a. The sample size is very large: 6448

b. \(H_0\): Attitude toward building new nuclear power plants is independent of Country

\(H_a\): Attitude toward building new nuclear power plants is not independent of Country

### Observed Frequency \((f_{ij})\)

<table>
<thead>
<tr>
<th>Country</th>
<th>G.B.</th>
<th>France</th>
<th>Italy</th>
<th>Spain</th>
<th>Ger.</th>
<th>U.S.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly favor</td>
<td>141</td>
<td>161</td>
<td>298</td>
<td>133</td>
<td>128</td>
<td>204</td>
<td>1065</td>
</tr>
<tr>
<td>Favor</td>
<td>348</td>
<td>366</td>
<td>309</td>
<td>222</td>
<td>272</td>
<td>326</td>
<td>1843</td>
</tr>
<tr>
<td>Oppose</td>
<td>381</td>
<td>334</td>
<td>219</td>
<td>311</td>
<td>322</td>
<td>316</td>
<td>1883</td>
</tr>
<tr>
<td>Strongly Oppose</td>
<td>217</td>
<td>215</td>
<td>219</td>
<td>443</td>
<td>389</td>
<td>174</td>
<td>1657</td>
</tr>
<tr>
<td>Total</td>
<td>1087</td>
<td>1076</td>
<td>1045</td>
<td>1109</td>
<td>1111</td>
<td>1020</td>
<td>6448</td>
</tr>
</tbody>
</table>

### Expected Frequency \((e_{ij})\)

\[ e_{ij} = \frac{(Row_{Total})(Column_{Total})}{SampleSize} \]

<table>
<thead>
<tr>
<th>Country</th>
<th>G.B.</th>
<th>France</th>
<th>Italy</th>
<th>Spain</th>
<th>Ger.</th>
<th>U.S.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly favor</td>
<td>179.5</td>
<td>177.7</td>
<td>172.6</td>
<td>183.2</td>
<td>183.5</td>
<td>168.5</td>
<td>1065</td>
</tr>
<tr>
<td>Favor</td>
<td>310.7</td>
<td>307.5</td>
<td>298.7</td>
<td>317.0</td>
<td>317.6</td>
<td>291.5</td>
<td>1843</td>
</tr>
<tr>
<td>Oppose</td>
<td>317.4</td>
<td>314.2</td>
<td>305.2</td>
<td>323.9</td>
<td>324.4</td>
<td>297.9</td>
<td>1883</td>
</tr>
<tr>
<td>Strongly Oppose</td>
<td>279.3</td>
<td>276.5</td>
<td>268.5</td>
<td>285.0</td>
<td>285.5</td>
<td>262.1</td>
<td>1657</td>
</tr>
<tr>
<td>Total</td>
<td>1087</td>
<td>1076</td>
<td>1045</td>
<td>1109</td>
<td>1111</td>
<td>1020</td>
<td>6448</td>
</tr>
</tbody>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
Tests of Goodness of Fit and Independence

Chi Square Calculations \( \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \)

<table>
<thead>
<tr>
<th>Country</th>
<th>G.B.</th>
<th>France</th>
<th>Italy</th>
<th>Spain</th>
<th>Ger.</th>
<th>U.S.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly favor</td>
<td>8.3</td>
<td>1.6</td>
<td>91.1</td>
<td>13.7</td>
<td>16.8</td>
<td>7.5</td>
<td>139.0</td>
</tr>
<tr>
<td>Favor</td>
<td>4.5</td>
<td>11.1</td>
<td>0.4</td>
<td>28.5</td>
<td>6.5</td>
<td>4.1</td>
<td>55.0</td>
</tr>
<tr>
<td>Oppose</td>
<td>12.7</td>
<td>1.2</td>
<td>24.3</td>
<td>0.5</td>
<td>0.0</td>
<td>1.1</td>
<td>39.9</td>
</tr>
<tr>
<td>Strongly Oppose</td>
<td>13.9</td>
<td>13.7</td>
<td>9.1</td>
<td>87.6</td>
<td>37.5</td>
<td>29.6</td>
<td>191.5</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 425.4 \]
\[ \chi^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \]

Degrees of freedom = \((r - 1)(c - 1) = (4-1)(6-1) = 15\)

Using the \( \chi^2 \) table with \( df = 15 \), \( \chi^2 = 425.4 \), the \( p \)-value is less than .005

Using Excel, the \( p \)-value corresponding to \( \chi^2 = 425.4; =CHISQ.DIST.RT(425.4,15) = .0000 \)

\( p \)-value ≤ .05, reject \( H_0 \). The attitude toward building new nuclear power plants is not independent of the country. Attitudes can be expected to vary with the country.

c. Use column percentages from the observed frequencies table to help answer this question.

Column percentages = Response frequency/Column Totals: For example, 141/1087 = 13.0%

<table>
<thead>
<tr>
<th>Country</th>
<th>G.B.</th>
<th>France</th>
<th>Italy</th>
<th>Spain</th>
<th>Ger.</th>
<th>U.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly favor</td>
<td>13.0%</td>
<td>15.0%</td>
<td>28.5%</td>
<td>12.0%</td>
<td>11.5%</td>
<td>20.0%</td>
</tr>
<tr>
<td>Favor</td>
<td>32.0%</td>
<td>34.0%</td>
<td>29.6%</td>
<td>20.0%</td>
<td>24.5%</td>
<td>32.0%</td>
</tr>
<tr>
<td>Oppose</td>
<td>35.1%</td>
<td>31.0%</td>
<td>21.0%</td>
<td>28.0%</td>
<td>29.0%</td>
<td>31.0%</td>
</tr>
<tr>
<td>Strongly Oppose</td>
<td>20.0%</td>
<td>20.0%</td>
<td>21.0%</td>
<td>39.9%</td>
<td>35.0%</td>
<td>17.0%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Adding together the percentages of respondents who “Strongly favor” and those who “Favor”, we find the following: Great Britain 45%, France 49%, Italy 58%, Spain 32%, Germany 36% and United States 52%. Italy shows the most support for nuclear power plants with 58% in favor. Spain shows the least support with only 32% in favor. Only Italy and the United States show more than 50% of the respondents in favor of building new nuclear power plants.

16. \( H_0 \): Movie reviews are independent of show host

\( H_a \): Movie reviews are not independent of show host

Expected Frequencies:

\[ e_{ij} = \frac{(Row_{Total})(Column_{Total})}{SampleSize} \]

\[ e_{11} = 11.81 \]
\[ e_{12} = 8.44 \]
\[ e_{13} = 24.75 \]
\[ e_{21} = 8.40 \]
\[ e_{22} = 6.00 \]
\[ e_{23} = 17.60 \]
\[ e_{31} = 21.79 \]
\[ e_{32} = 15.56 \]
\[ e_{33} = 45.65 \]

<table>
<thead>
<tr>
<th>Host A</th>
<th>Host B</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
<th>Chi Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con</td>
<td>Con</td>
<td>24</td>
<td>11.81</td>
<td>12.57</td>
</tr>
<tr>
<td>Con</td>
<td>Mixed</td>
<td>8</td>
<td>8.44</td>
<td>.02</td>
</tr>
</tbody>
</table>
$\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}$

Degrees of freedom $= (r - 1)(c - 1) = (3-1)(3-1) = 4$

Using the $\chi^2$ table with $df = 4$, $\chi^2 = 45.36$ shows the $p$-value is less than .005.

Using Excel, the $p$-value corresponding to $\chi^2 = 45.36 = \text{CHISQ.DIST.RT}(45.36,4) = .0000$.

$p$-value ≤ .01, reject $H_0$. Conclude that the ratings of the two hosts are not independent. The host responses are more similar than different and they tend to agree or be close in their ratings.

17. $H_0$: $p_1 = p_2 = p_3$

$H_a$: Not all population proportions are equal

Observed Frequencies ($f_{ij}$)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>150</td>
<td>150</td>
<td>96</td>
<td>396</td>
</tr>
<tr>
<td>No</td>
<td>100</td>
<td>150</td>
<td>104</td>
<td>354</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>300</td>
<td>200</td>
<td>750</td>
</tr>
</tbody>
</table>

Expected Frequencies ($e_{ij}$)  

$e_{ij} = \frac{(Row_{Total})(Column_{Total})}{SampleSize}$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>132.0</td>
<td>158.4</td>
<td>105.6</td>
<td>396</td>
</tr>
<tr>
<td>No</td>
<td>118.0</td>
<td>141.6</td>
<td>94.4</td>
<td>354</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>300</td>
<td>200</td>
<td>750</td>
</tr>
</tbody>
</table>

Chi Square Calculations ($f_{ij} - e_{ij})^2 / e_{ij}$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>2.45</td>
<td>.45</td>
<td>.87</td>
<td>3.77</td>
</tr>
<tr>
<td>No</td>
<td>2.75</td>
<td>.50</td>
<td>.98</td>
<td>4.22</td>
</tr>
</tbody>
</table>

$\chi^2 = 7.99$

Degrees of freedom $= k - 1 = (3 - 1) = 2$

Using the $\chi^2$ table with $df = 2$, $\chi^2 = 7.99$ shows the $p$-value is between .01 and .025

Using Excel, the $p$-value corresponding to $\chi^2 = 7.99 = \text{CHISQ.DIST.RT}(7.99,2) = .0184$
Tests of Goodness of Fit and Independence

$p$-value $\leq .05$, reject $H_0$. Conclude not all population proportions are equal.

18. a. $\bar{p}_1 = \frac{150}{250} = .60$

$\bar{p}_2 = \frac{150}{300} = .50$

$\bar{p}_3 = \frac{96}{200} = .48$

b. Multiple comparisons

$df = k - 1 = 3 - 1 = 2 \quad \chi^2_{.05} = 5.991$

Showing example calculation for 1 vs. 2

$$CV_{12} = \sqrt{\frac{2}{\chi^2} \left( \frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2} \right)} = \sqrt{\frac{60 (1 - .60)}{250} + \frac{.50 (1 - .50)}{300}} = .1037$$

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$p_1$</th>
<th>$p_2$</th>
<th>Difference</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>Critical Value</th>
<th>Significant Diff $&gt; CV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vs. 2</td>
<td>.60</td>
<td>.50</td>
<td>.10</td>
<td>250</td>
<td>300</td>
<td>.1037</td>
<td>No</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>.60</td>
<td>.48</td>
<td>.12</td>
<td>250</td>
<td>200</td>
<td>.1150</td>
<td>Yes</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>.50</td>
<td>.48</td>
<td>.02</td>
<td>300</td>
<td>200</td>
<td>.1117</td>
<td>No</td>
</tr>
</tbody>
</table>

Only one comparison is significant, 1 vs. 3. The others are not significant. We can conclude that the population proportions differ for populations 1 and 3.

20. a. $H_0$: $p_1 = p_2 = p_3$

$H_a$: Not all population proportions are equal

b. Observed Frequencies ($f_{ij}$)

<table>
<thead>
<tr>
<th>Component</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defective</td>
<td>15</td>
<td>20</td>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>Good</td>
<td>485</td>
<td>480</td>
<td>460</td>
<td>1425</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>1500</td>
</tr>
</tbody>
</table>

Expected Frequencies ($e_{ij}$)

$$e_{ij} = \frac{(Row_i \cdot Total) \cdot (Column_j \cdot Total)}{SampleSize}$$

<table>
<thead>
<tr>
<th>Component</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defective</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>Good</td>
<td>475</td>
<td>475</td>
<td>475</td>
<td>1425</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>1500</td>
</tr>
</tbody>
</table>

Chi Square Calculations ($f_{ij} - e_{ij})^2 / e_{ij}$

<table>
<thead>
<tr>
<th>Component</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
</table>
Defective & 4.00 & 1.00 & 9.00 & 14.00 \\
Good & .21 & .05 & .47 & 0.74 \\
\[\chi^2 = 14.74\]

Degrees of freedom \(k - 1 = (3 - 1) = 2\)

Using the \(\chi^2\) table with \(df = 2\), \(\chi^2 = 14.74\) shows the \(p\)-value is less than .01

Using Excel, the \(p\)-value corresponding to \(\chi^2 = 14.74\) = CHISQ.DIST.RT(14.74,2) = .0006

\(p\)-value \(\leq .05\), reject \(H_0\). Conclude that the three suppliers do not provide equal proportions of defective components.

c. \(\bar{p}_1 = \frac{15}{500} = .03\)
\(\bar{p}_2 = \frac{20}{500} = .04\)
\(\bar{p}_3 = \frac{40}{500} = .08\)

Multiple comparisons

For Supplier A vs. Supplier B

\(df = k - 1 = 3 - 1 = 2 \chi^2_{05} = 5.991\)

Showing example calculation for A vs B

\[
CV_{ij} = \sqrt{\chi^2 \frac{\bar{p}_i(1 - \bar{p}_i) + \bar{p}_j(1 - \bar{p}_j)}{n_i n_j}} = \sqrt{5.991 \left( \frac{0.03(1 - 0.03)}{500} + \frac{0.04(1 - 0.04)}{500} \right)} = .0284
\]

<table>
<thead>
<tr>
<th>Comparison</th>
<th>(p_1)</th>
<th>(p_2)</th>
<th>Difference</th>
<th>Critical Value</th>
<th>Significant</th>
<th>Diff &gt; CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>A vs. B</td>
<td>.03</td>
<td>.04</td>
<td>.01</td>
<td>500</td>
<td>500</td>
<td>.0284</td>
</tr>
<tr>
<td>A vs. C</td>
<td>.03</td>
<td>.08</td>
<td>.05</td>
<td>500</td>
<td>500</td>
<td>.0351</td>
</tr>
<tr>
<td>B vs. C</td>
<td>.04</td>
<td>.08</td>
<td>.04</td>
<td>500</td>
<td>500</td>
<td>.0366</td>
</tr>
</tbody>
</table>

Supplier A and supplier B are both significantly different from supplier C. Supplier C can be eliminated on the basis of a significantly higher proportion of defective components. Since suppliers A and supplier B are not significantly different in terms of the proportion defective components, both of these suppliers should remain candidates for use by Benson.

22. a. \(\bar{p}_1 = \frac{35}{250} = .14\) 14% error rate
\(\bar{p}_2 = \frac{27}{300} = .09\) 9% error rate

b. \(H_0: p_1 - p_2 = 0\)
\(H_a: p_1 - p_2 \neq 0\)

Observed Frequencies \((f_0)\)
Tests of Goodness of Fit and Independence

<table>
<thead>
<tr>
<th>Return</th>
<th>Office 1</th>
<th>Office 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>35</td>
<td>27</td>
<td>62</td>
</tr>
<tr>
<td>Correct</td>
<td>215</td>
<td>273</td>
<td>488</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>300</td>
<td>550</td>
</tr>
</tbody>
</table>

Expected Frequencies \( (e_{ij}) \)

\[
e_{ij} = \frac{(\text{Row Total})(\text{Column Total})}{\text{Sample Size}}
\]

<table>
<thead>
<tr>
<th>Return</th>
<th>Office 1</th>
<th>Office 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>28.18</td>
<td>33.82</td>
<td>62</td>
</tr>
<tr>
<td>Correct</td>
<td>221.82</td>
<td>266.18</td>
<td>488</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>300</td>
<td>550</td>
</tr>
</tbody>
</table>

Chi Square Calculations \( (f_{ij} - e_{ij})^2 / e_{ij} \)

<table>
<thead>
<tr>
<th>Return</th>
<th>Office 1</th>
<th>Office 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>1.65</td>
<td>1.37</td>
<td>3.02</td>
</tr>
<tr>
<td>Correct</td>
<td>.21</td>
<td>.17</td>
<td>.38</td>
</tr>
</tbody>
</table>

\[
\chi^2 = 3.41
\]

\[
\chi^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}}
\]

\( df = k - 1 = (2 - 1) = 1 \)

Using the \( \chi^2 \) table with \( df = 1, \chi^2 = 3.41 \) shows the \( p \)-value is between .05 and .10

Using Excel, the \( p \)-value corresponding to \( \chi^2 = 3.41 = \text{CHISQ.DIST.RT}(3.41,1) = .0648 \)

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0649

\( p \)-value \( \leq .10 \), reject \( H_0 \). Conclude that the two offices do not have the same population proportion error rates.

c. With two populations, a chi–square test for equal population proportions has 1 degree of freedom. In this case the test statistic \( \chi^2 \) is always equal to \( z^2 \). This relationship between the two test statistics always provides the same \( p \)-value and the same conclusion when the null hypothesis involves equal population proportions. However, the use of the \( z \) test statistic provides options for one–tailed hypothesis tests about two population proportions while the chi–square test is limited a two–tailed hypothesis tests about the equality of the two population proportions.

24. \( H_0 \): The distribution of defects is the same for all suppliers

\( H_a \): The distribution of defects is not the same all suppliers

Observed Frequencies \( (f_{ij}) \)

<table>
<thead>
<tr>
<th>Part Tested</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor Defect</td>
<td>15</td>
<td>13</td>
<td>21</td>
<td>49</td>
</tr>
<tr>
<td>Major Defect</td>
<td>5</td>
<td>11</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Good</td>
<td>130</td>
<td>126</td>
<td>124</td>
<td>380</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>450</td>
</tr>
</tbody>
</table>
Chapter 12

Expected Frequencies \((e_{ij})\)

\[ e_{ij} = \frac{(Row_i \cdot Total_j)(Column_i \cdot Total_j)}{Sample Size} \]

<table>
<thead>
<tr>
<th>Part Tested</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor Defect</td>
<td>16.33</td>
<td>16.33</td>
<td>16.33</td>
<td>49</td>
</tr>
<tr>
<td>Major Defect</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>21</td>
</tr>
<tr>
<td>Good</td>
<td>126.67</td>
<td>126.67</td>
<td>126.67</td>
<td>380</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>450</td>
</tr>
</tbody>
</table>

Chi Square Calculations \((f_{ij} - e_{ij})^2 / e_{ij}\)

<table>
<thead>
<tr>
<th>Part Tested</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor Defect</td>
<td>.11</td>
<td>.68</td>
<td>1.33</td>
<td>2.12</td>
</tr>
<tr>
<td>Major Defect</td>
<td>.57</td>
<td>2.29</td>
<td>.57</td>
<td>3.43</td>
</tr>
<tr>
<td>Good</td>
<td>.09</td>
<td>.00</td>
<td>.06</td>
<td>.15</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \]

Degrees of freedom = \((r - 1)(k - 1) = (3 - 1)(3 - 1) = 4\)

Using the \(\chi^2\) table with \(df = 4\), \(\chi^2 = 5.70\) shows the \(p\)-value is greater than .10

Using Excel, the \(p\)-value corresponding to \(\chi^2 = 5.70\) = CHISQ.DIST.RT(5.7,4) = .2227

Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .2228

\(p\)-value > .05, do not reject \(H_0\). Conclude that we are unable to reject the hypothesis that the population distribution of defects is the same for all three suppliers. There is no evidence that quality of parts from one supplier is better than either of the others two suppliers.

26. a. \(H_0\): Order pattern probabilities for Dayton are consistent with established Bistro 65 restaurants with Pasta .4, Steak&Chops .1, Seafood .2 and Other .3

\(H_a\): Order pattern probabilities for Dayton are not the same as established Bistro 65 restaurants

Shown below is the frequency distribution for sales at the new Dayton restaurant.

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasta</td>
<td>70</td>
</tr>
<tr>
<td>Steak &amp; Chops</td>
<td>30</td>
</tr>
<tr>
<td>Seafood</td>
<td>50</td>
</tr>
<tr>
<td>Other</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
</tr>
</tbody>
</table>

If the order pattern for the new restaurant in Dayton is the same as the historical pattern for the established Bistro 65 restaurants, the expected number of orders for each category would be as follows:

<table>
<thead>
<tr>
<th>Category</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pasta</td>
<td>.4(200) = 80</td>
</tr>
<tr>
<td>Steak &amp; Chops</td>
<td>.1(200) = 20</td>
</tr>
</tbody>
</table>
Seafood \( \frac{2(200)}{} = 40 \)
Other \( \frac{3(200)}{} = 60 \)

Total 200

<table>
<thead>
<tr>
<th>Category</th>
<th>Hypothesized Proportion (p)</th>
<th>Observed Frequency ( (f_i) )</th>
<th>Expected Frequency ( (e_i) = n*p )</th>
<th>Chi Square ( \frac{(f_i - e_i)^2}{e_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank</td>
<td>.4</td>
<td>70</td>
<td>80</td>
<td>1.25</td>
</tr>
<tr>
<td>Cable</td>
<td>.1</td>
<td>30</td>
<td>20</td>
<td>5.00</td>
</tr>
<tr>
<td>Car</td>
<td>.2</td>
<td>50</td>
<td>40</td>
<td>2.50</td>
</tr>
<tr>
<td>Collection</td>
<td>.3</td>
<td>50</td>
<td>60</td>
<td>1.67</td>
</tr>
<tr>
<td>Total</td>
<td>200 = n</td>
<td></td>
<td></td>
<td>( \chi^2 = 10.42 )</td>
</tr>
</tbody>
</table>

\( k - 1 = 4 - 1 = 3 \) degrees of freedom

Using the \( \chi^2 \) table with \( df = 3 \), \( \chi^2 = 10.42 \) shows the \( p \)-value is between .01 and .025.
Using Excel, the \( p \)-value corresponding to \( \chi^2 = 10.42 \) = CHISQ.DIST.RT(10.42,3) = .0153

\( p \)-value \leq .05, reject \( H_0 \). We reject the hypothesis that the order pattern for Dayton is the same as the order pattern of established Bistro 65 restaurants.

Similarly, using Excel’s CHISQ.TEST function with the above observed and expected frequencies the \( p \)-value associated with the chi-square test for goodness of fit is .0153. Because the \( p \)-value = .0153 \( \leq \alpha = .05 \), we reject the hypothesis the order pattern for the new restaurant in Dayton is the same as the order pattern for the established Bistro 65 restaurants.

b. The side-by-side bar chart below compares the purchase preference probabilities for the new restaurant in Dayton and the established or “old” Bistro 65 restaurants. We see that the new restaurant sells a larger percentage of Steak & Chops and Seafood, but a lower percentage of Pasta and Other foods.
28. a.  

\( H_0: \) The preferred pace of life is independent of gender  
\( H_a: \) The preferred pace of life is not independent of gender

**Observed Frequency \((f_{ij})\)**

<table>
<thead>
<tr>
<th>Preferred Pace of Life</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slower</td>
<td>230</td>
<td>218</td>
<td>448</td>
</tr>
<tr>
<td>No Preference</td>
<td>20</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>Faster</td>
<td>90</td>
<td>48</td>
<td>138</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>340</td>
<td>290</td>
<td>630</td>
</tr>
</tbody>
</table>

**Expected Frequency \((e_{ij})\)**

\[
e_{ij} = \frac{(Row_i \times Total)(Column_j \times Total)}{SampleSize}
\]

<table>
<thead>
<tr>
<th>Preferred Pace of Life</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slower</td>
<td>241.78</td>
<td>206.22</td>
<td>448</td>
</tr>
<tr>
<td>No Preference</td>
<td>23.75</td>
<td>20.25</td>
<td>44</td>
</tr>
<tr>
<td>Faster</td>
<td>74.48</td>
<td>63.52</td>
<td>138</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>340</td>
<td>290</td>
<td>630</td>
</tr>
</tbody>
</table>

**Chi Square Calculations \((f_{ij} - e_{ij})^2 / e_{ij}\)**

<table>
<thead>
<tr>
<th>Preferred Pace of Life</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slower</td>
<td>.57</td>
<td>.67</td>
<td>1.25</td>
</tr>
<tr>
<td>No Preference</td>
<td>.59</td>
<td>.69</td>
<td>1.28</td>
</tr>
<tr>
<td>Faster</td>
<td>3.24</td>
<td>3.79</td>
<td>7.03</td>
</tr>
</tbody>
</table>

\[
\chi^2 = 9.56
\]

Degrees of freedom \(= (r - 1)(c - 1) = (3 - 1)(2 - 1) = 2\)

Using the \(\chi^2\) table with \(df = 2\), \(\chi^2 = 9.56\) shows the \(p\)-value is between .005 and .01.

Using Excel, the \(p\)-value corresponding to \(\chi^2 = 9.56\) = CHISQ.DIST.RT(9.56,2) = .0084.

\(p\)-value \(\leq .05\), reject \(H_0\). The preferred pace of life is not independent of gender. Thus, we expect men and women differ with respect to the preferred pace of life.

b. Percentage responses for each gender

Column percentages = Gender & Pace Frequency / Gender Total; for example: \(230/340 = 67.65\%\)

<table>
<thead>
<tr>
<th>Preferred Pace of Life</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slower</td>
<td>67.65%</td>
<td>75.17%</td>
</tr>
<tr>
<td>No Preference</td>
<td>5.88%</td>
<td>8.28%</td>
</tr>
<tr>
<td>Faster</td>
<td>26.47%</td>
<td>16.55%</td>
</tr>
</tbody>
</table>
The highest percentages are for a slower pace of life by both men and women. However, 75.17% of women prefer a slower pace compared to 67.65% of men and 26.47% of men prefer a faster pace compared to 16.55% of women. More women prefer a slower pace while more men prefer a faster pace.

30. \( H_0: \) Emergency calls within each county are independent of the day of week

\( H_a: \) Emergency calls within each county are not independent of the day of week

| Observed Frequencies \( (f_{ij}) \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Day of Week     | Sun             | Mon             | Tues            | Wed             | Thu             | Fri             | Sat             | Total           |
| Urban           | 61             | 48             | 50             | 55             | 63             | 73             | 43             | 393             |
| Rural           | 7             | 9             | 16             | 13             | 9             | 14             | 10             | 78             |
| Total           | 68             | 57             | 66             | 68             | 72             | 87             | 53             | 471             |

| Expected Frequencies \( (e_{ij}) \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( e_{ij} = \frac{(Row_Total)(Column_Total)}{SampleSize} \) |

| Day of Week     | Sun             | Mon             | Tues            | Wed             | Thu             | Fri             | Sat             | Total           |
| Urban           | 56.74           | 47.56           | 55.07           | 56.74           | 60.08           | 72.59           | 44.22           | 393             |
| Rural           | 11.26           | 9.44            | 10.93           | 11.26           | 11.92           | 14.41           | 8.78            | 78              |
| Total           | 68             | 57             | 66             | 68             | 72             | 87             | 53             | 471             |

| Chi Square \( (f_{ij} - e_{ij})^2/e_{ij} \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Day of Week     | Sun             | Mon             | Tues            | Wed             | Thu             | Fri             | Sat             | Total           |
| Urban           | .32             | .00             | .47             | .05             | .14             | .00             | .03             | 1.02            |
| Rural           | 1.61            | .02             | 2.35            | .27             | .72             | .01             | .17             | 5.15            |

\( \chi^2 = 6.17 \)

Degrees of freedom \( = (r - 1)(c - 1) = (2 - 1)(7 - 1) = 6 \)

Using the \( \chi^2 \) table with \( df = 6 \), \( \chi^2 = 6.17 \) shows the \( p \)-value is greater than .10.

Using Excel, the \( p \)-value corresponding to \( \chi^2 = 6.17 = \text{CHISQ.DIST.RT}(6.17,6) = .4044 \).

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .4039

\( p \)-value > .05, do not reject \( H_0 \). The assumption of independence cannot be rejected. The county with the emergency call does not vary or depend upon the day of the week.
32. a. \( p_1 = \frac{44}{500} = .088 \)

\( p_2 = \frac{35}{300} = .117 \)

\( p_3 = \frac{36}{400} = .090 \)

\( p_4 = \frac{34}{400} = .085 \)

Bridgeport 8.8%, Los Alamos 11.7%, Naples 9%, Washington DC 8.5%

b. \( H_0: p_1 = p_2 = p_3 = p_4 \)

\( H_a: \) Not all population proportions are equal

Observed Frequencies (\( f_{ij} \))

<table>
<thead>
<tr>
<th>Millionaire</th>
<th>Bridgeport</th>
<th>Los' Alamos</th>
<th>Naples</th>
<th>Washington</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>44</td>
<td>35</td>
<td>36</td>
<td>34</td>
<td>149</td>
</tr>
<tr>
<td>No</td>
<td>456</td>
<td>265</td>
<td>364</td>
<td>366</td>
<td>1451</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>300</td>
<td>400</td>
<td>400</td>
<td>1600</td>
</tr>
</tbody>
</table>

Expected Frequencies (\( e_{ij} \))

\[
e_{ij} = \frac{(\text{RowTotal})(\text{ColumnTotal})}{\text{SampleSize}}
\]

<table>
<thead>
<tr>
<th>Millionaire</th>
<th>Bridgeport</th>
<th>Los' Alamos</th>
<th>Naples</th>
<th>Washington</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>46.56</td>
<td>27.94</td>
<td>37.25</td>
<td>37.25</td>
<td>149</td>
</tr>
<tr>
<td>No</td>
<td>453.44</td>
<td>272.06</td>
<td>362.75</td>
<td>362.75</td>
<td>1451</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>300</td>
<td>400</td>
<td>400</td>
<td>1600</td>
</tr>
</tbody>
</table>

Chi Square Calculations \((f_{ij} - e_{ij})^2 / e_{ij}\)

<table>
<thead>
<tr>
<th>Millionaire</th>
<th>Bridgeport</th>
<th>Los' Alamos</th>
<th>Naples</th>
<th>Washington</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>.14</td>
<td>1.79</td>
<td>.04</td>
<td>.28</td>
<td>2.25</td>
</tr>
<tr>
<td>No</td>
<td>.01</td>
<td>.18</td>
<td>.00</td>
<td>.03</td>
<td>.23</td>
</tr>
</tbody>
</table>

\( \chi^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = 2.48 \)

\( \chi^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \)

Degrees of freedom = \( k - 1 = (4 - 1) = 3 \)

Using the \( \chi^2 \) table with \( df = 3 \), \( \chi^2 = 2.48 \) shows the \( p \)-value is greater than .10

Using Excel, the \( p \)-value corresponding to \( \chi^2 = 2.48 = \text{CHISQ.DIST.RT}(2.48,3) = .4789 \)

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .4783

\( p \)-value > .05, do not reject \( H_0 \). Cannot conclude that there is a difference among the population proportion of millionaires for these four cities.
Chapter 13
Experimental Design and Analysis of Variance

Solutions:

1. a. \( \bar{x} = \frac{(156 + 142 + 134)}{3} = 144 \)  
   Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means. 

   \[ SSTR = \sum_{j=1}^{k} n_j \left( \bar{x}_j - \bar{x} \right)^2 = 6(156 - 144)^2 + 6(142 - 144)^2 + 6(134 - 144)^2 = 1,488 \]

   b. \( \text{MSTR} = \frac{SSTR}{(k - 1)} = \frac{1488}{2} = 744 \)

   c. \( s_1^2 = 164.4 \quad s_2^2 = 131.2 \quad s_3^2 = 110.4 \)

   \[ SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 5(164.4) + 5(131.2) + 5(110.4) = 2030 \]

   d. \( \text{MSE} = \frac{SSE}{(n_T - k)} = \frac{2030}{(18 - 3)} = 135.3 \)

   e. Using data in problem, the Excel ANOVA (Single Factor) tool can be used to generate table (Note that in the Excel generated output, the Between Groups Variation is the Treatments and Within Groups Variation is Error.), or values can be filled in from parts a-d, adding \( F = \frac{\text{MSTR}}{\text{MSE}} = \frac{744}{135.3} = 5.5 \)

   Using Excel, the \( p \)-value corresponding to \( F = 5.5 \) = F.DIST.RT(5.5,2,15) = .0162

   ![](source of variation table)

   f. \( H_0: \mu_1 = \mu_2 = \mu_3 \) 

   \( H_a: \) Not all the treatment population means are equal 

   \[ F = \frac{\text{MSTR}}{\text{MSE}} = \frac{744}{135.3} = 5.5 \]

   Using \( F \) table (2 degrees of freedom numerator and 15 denominator), \( p \)-value is between .01 and .025 

   Using Excel, the \( p \)-value corresponding to \( F = 5.5 \) = F.DIST.RT(5.5,2,15) = .0162.

   Because \( p \)-value \( \leq \alpha = .05 \), we reject the hypothesis that the means for the three treatments are equal.
2. Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F$ | $p$-value
--- | --- | --- | --- | --- | ---
Treatments | 300 | 4 | 75 | 14.07 | .0000
Error | 160 | 30 | 5.33 | |
Total | 460 | 34 | |

\[ \text{SSE} = \text{SST} - \text{SSTR} = 460 - 300 = 160 \]

Treatments degrees of freedom = $k-1 = 5-1 = 4$, where $k$ = the number of factors/treatments/samples being compared
Total observations = $n_T = 5*7 = 35$

Total df = $n_T-1 = 35-1 = 34$
Error df = $n_T-k = 35 - 5 = 30$

\[
\text{MSTR} = \frac{\text{SSTR}}{k-1} = \frac{300}{4} = 75 \\
\text{MSE} = \frac{\text{SSE}}{n_T-k} = \frac{160}{30} = 5.33 \\
\]

\[ F = \frac{\text{MSTR}}{\text{MSE}} = \frac{75}{5.33} = 14.07 \]

Using Excel, the $p$-value corresponding to $F = 14.07 = \text{F.DIST.RT}(14.07,4,30) = .0000$.

4. $H_0$: $\mu_1 = \mu_2 = \mu_3$
$H_a$: Not all the treatment population means are equal

Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F$ | $p$-value
--- | --- | --- | --- | --- | ---
Treatments | 150 | 2 | 75 | 4.80 | .0233
Error | 250 | 16 | 15.625 | |
Total | 400 | 18 | |

\[ \text{SSE} = \text{SST} - \text{SSTR} = 400 - 150 = 250 \]

Treatments degrees of freedom = $k-1 = 3-1 = 2$, where $k$ = the number of factors/treatments/samples being compared
Total observations = $n_T = 19$

Total df = $n_T-1 = 19-1 = 18$
Error df = $n_T-k = 19 - 3 = 16$

\[
\text{MSTR} = \frac{\text{SSTR}}{k-1} = \frac{150}{2} = 75 \\
\text{MSE} = \frac{\text{SSE}}{n_T-k} = \frac{250}{16} = 15.625 \\
\]

\[ F = \frac{\text{MSTR}}{\text{MSE}} = \frac{75}{15.625} = 4.8 \]

Using $F$ table (2 degrees of freedom numerator and 16 denominator), $p$-value is between .01 and .025
Using Excel, the $p$-value corresponding to $F = 4.80 = \text{F.DIST.RT}(4.8,2,16) = .0233$.

Because $p$-value $\leq \alpha = .05$, we reject the null hypothesis that the means of the three treatments are equal.
6. \(H_0: \mu_1 = \mu_2 = \mu_3\)
\(H_a: \) Not all the treatment population means are equal

Using Exer6 datafile, the Excel Single Factor ANOVA Output follows:

**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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1617.857</td>
<td>2</td>
<td>808.9286</td>
<td>5.8449</td>
<td>0.0083</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3460</td>
<td>25</td>
<td>138.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5077.857</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>952</td>
<td>119</td>
<td>146.857</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>1070</td>
<td>107</td>
<td>96.4444</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>1000</td>
<td>100</td>
<td>173.778</td>
</tr>
</tbody>
</table>

\[ \bar{x} = \frac{\sum_{j=1}^{k} n_j \bar{x}_j}{n_T} \]

\[ SSTR = \sum_{j=1}^{k} n_j (-\bar{x})^2 = 8(119 - 107.93)^2 + 10(107 - 107.93)^2 + 10(100 - 107.93)^2 = 1617.86 \]

\[ MSTR = \frac{SSTR}{k - 1} = 1617.86 / 2 = 808.93 \]

\[ SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 7(146.86) + 9(96.44) + 9(173.78) = 3,460 \]

\[ MSE = \frac{SSE}{n_T - k} = 3,460 / (28 - 3) = 138.4 \]

\[ F = \frac{MSTR}{MSE} = 808.93 / 138.4 = 5.84 \]

Using \(F\) table (2 degrees of freedom numerator and 25 denominator), \(p\)-value is less than .01

Using Excel, the \(p\)-value corresponding to \(F = 5.84\) = F.DIST.RT(5.84,2,25) = .0083.

Because \(p\)-value \(\leq \alpha = .05\), we reject the null hypothesis that the means of the three treatments are equal.
8.  

\[ H_0: \mu_1 = \mu_2 = \mu_3 \]

\[ H_a: \text{Not all the treatment population means are equal} \]

\[
\overline{x} = (79 + 74 + 66)/3 = 73
\]

Note: When the samplesizes are the same, the overall sample mean is an average of the individual sample means.

\[
\text{SSTR} = \sum_{j=1}^{k} n_j (\overline{x}_j - \overline{x})^2 = 6(79 - 73)^2 + 6(74 - 73)^2 + 6(66 - 73)^2 = 516
\]

\[
\text{MSTR} = \frac{\text{SSTR}}{(k - 1)} = \frac{516}{2} = 258
\]

\[
s_1^2 = 34 \quad s_2^2 = 20 \quad s_3^2 = 32
\]

\[
\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 5(34) + 5(20) + 5(32) = 430
\]

\[
\text{MSE} = \frac{\text{SSE}}{(n_T - k)} = \frac{430}{(18 - 3)} = 28.67
\]

\[
F = \frac{\text{MSTR}}{\text{MSE}} = \frac{258}{28.67} = 9.00
\]

Using data in NCP datafile, the Excel ANOVA (Single Factor) tool can be used to generate table (Note that in the Excel generated output, the Between Groups Variation is the Treatments and Within Groups Variation is Error.), or values can be filled in from calculations above

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>516</td>
<td>2</td>
<td>258</td>
<td>9.00</td>
<td>.0027</td>
</tr>
<tr>
<td>Error</td>
<td>430</td>
<td>15</td>
<td>28.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>946</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using \( F \) table (2 degrees of freedom numerator and 15 denominator), \( p \)-value is less than \( .01 \). Using Excel the \( p \)-value corresponding to \( F = 9.00 = \text{F.DIST.RT(9,2,15)} = .0027 \).

Because \( p \)-value \( \leq \alpha = .05 \), we reject the null hypothesis that the means for the three plants are equal. In other words, analysis of variance supports the conclusion that the population mean examination score at the three NCP plants are not equal.

10. \n
\[ H_0: \mu_1 = \mu_2 = \mu_3 \]

\[ H_a: \text{Not all the treatment population means are equal} \]

Using AudJudg datafile, the Excel Single Factor ANOVA Output follows:

**Anova: Single Factor**

**SUMMARY**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>7</td>
<td>119</td>
<td>17</td>
<td>5.01</td>
</tr>
<tr>
<td>Indirect</td>
<td>7</td>
<td>142.8</td>
<td>20.4</td>
<td>6.256667</td>
</tr>
<tr>
<td>Combination</td>
<td>7</td>
<td>175</td>
<td>25</td>
<td>4.01</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>225.68</td>
<td>2</td>
<td>112.84</td>
<td>22.15928</td>
<td>.000014</td>
</tr>
</tbody>
</table>
Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

<table>
<thead>
<tr>
<th></th>
<th>Direct Experience</th>
<th>Indirect Experience</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean</td>
<td>17.0</td>
<td>20.4</td>
<td>25.0</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>5.01</td>
<td>6.2567</td>
<td>4.01</td>
</tr>
</tbody>
</table>

\[
\bar{x} = \frac{(17 + 20.4 + 25)}{3} = 20.8 \quad \text{Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.}
\]

\[
\text{SSTR} = \sum_{j=1}^{k} n_j \left( \bar{x}_j - \bar{x} \right)^2 = 7(17 - 20.8)^2 + 7(20.4 - 20.8)^2 + 7(25 - 20.8)^2 = 225.68
\]

\[
\text{MSTR} = \frac{\text{SSTR}}{k - 1} = \frac{225.68}{2} = 112.84
\]

\[
\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 6(5.01) + 6(6.2567) + 6(4.01) = 91.66
\]

\[
\text{MSE} = \frac{\text{SSE}}{n_T - k} = \frac{91.66}{(21 - 3)} = 5.092
\]

\[
F = \frac{\text{MSTR}}{\text{MSE}} = \frac{112.84}{5.092} = 22.16
\]

Using \(F\) table (2 degrees of freedom numerator and 18 denominator), \(p\)-value is less than .01

Using Excel the \(p\)-value corresponding to \(F = 22.16\) = F.DIST.RT(22.16,2,18) = .0000.

Because \(p\)-value \(\leq \alpha = .05\), we reject the null hypothesis that the means for the three groups are equal.

\[H_0: \mu_1 = \mu_2 = \mu_3\]

\[H_a: \text{Not all the treatment population means are equal}\]

Using GrandStrand datafile, the Excel Single Factor ANOVA Output follows:

Anova: Single Factor

**SUMMARY**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian</td>
<td>8</td>
<td>136</td>
<td>17</td>
<td>14.85714</td>
</tr>
<tr>
<td>Seafood</td>
<td>8</td>
<td>152</td>
<td>19</td>
<td>13.71429</td>
</tr>
<tr>
<td>Steakhouse</td>
<td>8</td>
<td>192</td>
<td>24</td>
<td>14</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>208</td>
<td>2</td>
<td>104</td>
<td>7.328859</td>
<td>0.003852</td>
</tr>
<tr>
<td>Within Groups</td>
<td>298</td>
<td>21</td>
<td>14.19048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>506</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.
Chapter 13

Italian Seafood Steakhouse

<table>
<thead>
<tr>
<th>Sample Mean</th>
<th>Italian</th>
<th>Seafood</th>
<th>Steakhouse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>14.857</td>
<td>13.714</td>
<td>14.000</td>
</tr>
</tbody>
</table>

\[ \bar{x} = \frac{(17 + 19 + 24)}{3} = 20 \]  
Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.

\[ \text{SSTR} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 8(17 - 20)^2 + 8(19 - 20)^2 + 8(24 - 20)^2 = 208 \]

\[ \text{MSTR} = \frac{\text{SSTR}}{(k - 1)} = \frac{208}{2} = 104 \]

\[ \text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 7(14.857) + 7(13.714) + 7(14.000) = 298 \]

\[ \text{MSE} = \frac{\text{SSE}}{(n_T - k)} = \frac{298}{(24 - 3)} = 14.19 \]

\[ F = \frac{\text{MSTR}}{\text{MSE}} = \frac{104}{14.19} = 7.33 \]

Using the F table (2 degrees of freedom numerator and 21 denominator), the p-value is less than .01. Using Excel, the p-value corresponding to \( F = 7.33 = \text{F.DIST.RT}(7.33,2,21) = .0038 \).

Using unrounded F test statistic, the pvalue = .0039

Because \( p\)-value ≤ \( \alpha \) = .05, we reject the null hypothesis that the mean meal prices are the same for the three types of restaurants.

13. a. \( H_0: \mu_1 = \mu_2 = \mu_3 \)
\( H_a: \) Not all the treatment population means are equal

Using data given in problem, the Excel Single Factor ANOVA Output follows:

**Summary**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment A</td>
<td>5</td>
<td>150</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Treatment B</td>
<td>5</td>
<td>225</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>Treatment C</td>
<td>5</td>
<td>180</td>
<td>36</td>
<td>6.5</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>570</td>
<td>2</td>
<td>285</td>
<td>51.81818</td>
<td>0.000001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>66</td>
<td>12</td>
<td>5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>636</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

\[ \bar{x} = \frac{(30 + 45 + 36)}{3} = 37 \]  
Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.
SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 5(30 - 37)^2 + 5(45 - 37)^2 + 5(36 - 37)^2 = 570

MSTR = \frac{SSTR}{k - 1} = \frac{570}{2} = 285

SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 4(6) + 4(4) + 4(6.5) = 66

MSE = \frac{SSE}{n_T - k} = \frac{66}{15 - 3} = 5.5

F = \frac{MSTR}{MSE} = \frac{285}{5.5} = 51.82

Using F table (2 degrees of freedom numerator and 12 denominator), p-value is less than .01

Using Excel, the p-value corresponding to F = 51.82 = F.DIST.RT(51.82,2,12) = .0000.

Because p-value ≤ α = .05, we reject the null hypothesis that the means of the three populations are equal.

b. \( t_{0.05} \) for 12 df = 2.179

\[ \text{LSD} = t_{0.05} \sqrt{\text{MSE} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = t_{0.05} \sqrt{5.5 \left( \frac{1}{5} + \frac{1}{5} \right)} = 2.179 \sqrt{2.2} = 3.23 \]

\(|\bar{x}_1 - \bar{x}_2| = |30 - 45| = 15 > \text{LSD}; \text{significant difference}\

\(|\bar{x}_1 - \bar{x}_1| = |30 - 36| = 6 > \text{LSD}; \text{significant difference}\

\(|\bar{x}_2 - \bar{x}_1| = |45 - 36| = 9 > \text{LSD}; \text{significant difference}\

c. \( \bar{x}_1 - \bar{x}_2 \pm t_{0.005} \sqrt{\text{MSE} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \)

\((30 - 45) \pm 2.179 \sqrt{5.5 \left( \frac{1}{5} + \frac{1}{5} \right)} \)

-15 ± 3.23 = -18.23 to -11.77

14. a. \( H_0: \mu_1 = \mu_2 = \mu_3 \)

\( H_a: \text{Not all the treatment population means are equal} \)

Using data given in problem, the Excel Single Factor ANOVA Output follows:

<table>
<thead>
<tr>
<th></th>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>4</td>
<td>204</td>
<td>51</td>
<td>96.66667</td>
<td></td>
</tr>
<tr>
<td>Treatment 2</td>
<td>4</td>
<td>308</td>
<td>77</td>
<td>97.33333</td>
<td></td>
</tr>
<tr>
<td>Treatment 3</td>
<td>4</td>
<td>232</td>
<td>58</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>

ANNOVA
### Source of Variation Summary

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1448</td>
<td>2</td>
<td>724</td>
<td>7.869</td>
<td>0.010565</td>
</tr>
<tr>
<td>Within Groups</td>
<td>828</td>
<td>9</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2276</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

### Sample Data

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample Mean</th>
<th>Sample Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>96.67</td>
</tr>
<tr>
<td>2</td>
<td>77</td>
<td>97.34</td>
</tr>
<tr>
<td>3</td>
<td>58</td>
<td>81.99</td>
</tr>
</tbody>
</table>

\[ \bar{x} = \frac{(51 + 77 + 58)}{3} = 62 \]

Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.

\[ SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 4(51 - 62)^2 + 4(77 - 62)^2 + 4(58 - 62)^2 = 1,448 \]

\[ MSTR = \frac{SSTR}{(k - 1)} = \frac{1,448}{2} = 724 \]

\[ SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 3(96.67) + 3(97.34) + 3(81.99) = 828 \]

\[ MSE = \frac{SSE}{(n_T - k)} = \frac{828}{12 - 3} = 92 \]

\[ F = \frac{MSTR}{MSE} = \frac{724}{92} = 7.87 \]

Using \( F \) table (2 degrees of freedom numerator and 9 denominator), \( p \)-value is between .01 and .025

Using Excel, the \( p \)-value corresponding to \( F = 7.87 \) is \( F.DIST.RT(7.87,2,9) = .0106 \)

Because \( p \)-value \( \leq \alpha = .05 \), we reject the null hypothesis that the means of the three populations are equal.

b. \( t_{0.05} \) for 9 df = 2.262

\[ LSD = t_{0.05} \sqrt{MSE \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} = t_{0.05} \sqrt{92 \left( \frac{1}{3} + \frac{1}{3} \right)} = 2.262 \sqrt{46} = 15.34 \]

\[ |\bar{x}_1 - \bar{x}_2| = |51 - 77| = 26 > LSD; \text{ significant difference} \]

\[ |\bar{x}_1 - \bar{x}_3| = |51 - 58| = 7 < LSD; \text{ no significant difference} \]

\[ |\bar{x}_2 - \bar{x}_3| = |77 - 58| = 19 > LSD; \text{ significant difference} \]

15. a. \( H_0: \mu_1 = \mu_2 = \mu_3 \)

\( H_a: \) Not all the treatment population means are equal

Using data given in problem, the Excel Single Factor ANOVA Output follows:
Anova: Single Factor

### SUMMARY

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mfg 1</td>
<td>4</td>
<td>92</td>
<td>23</td>
<td>6.666667</td>
</tr>
<tr>
<td>Mfg 2</td>
<td>4</td>
<td>112</td>
<td>28</td>
<td>4.666667</td>
</tr>
<tr>
<td>Mfg 3</td>
<td>4</td>
<td>84</td>
<td>21</td>
<td>3.333333</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>104</td>
<td>2</td>
<td>52</td>
<td>10.636</td>
<td>0.00426</td>
</tr>
<tr>
<td>Within Groups</td>
<td>44</td>
<td>9</td>
<td>4.889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>148</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

Manufacturer 1 | Manufacturer 2 | Manufacturer 3
Sample Mean   | 23         | 28         | 21         |
Sample Variance | 6.667     | 4.667     | 3.333     |

\[ \bar{x} = \frac{(23 + 28 + 21)}{3} = 24 \]
Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.

\[ SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 4(23 - 24)^2 + 4(28 - 24)^2 + 4(21 - 24)^2 = 104 \]

\[ MSTR = \frac{SSTR}{(k - 1)} = \frac{104}{2} = 52 \]

\[ SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 3(6.667) + 3(4.667) + 3(3.33) = 44 \]

\[ MSE = \frac{SSE}{(n_T - k)} = \frac{44}{12 - 3} = 4.889 \]

\[ F = \frac{MSTR}{MSE} = \frac{52}{4.889} = 10.636 \]

Using \( F \) table (2 degrees of freedom numerator and 9 denominator), \( p \)-value is less than .01
Using Excel, the \( p \)-value corresponding to \( F = 10.636 \) = F.DIST.RT(10.636,2,9) = .0043

Because \( p \)-value ≤ \( \alpha \) = .05, we reject the null hypothesis that the mean time needed to mix a batch of material is the same for each manufacturer.

b. \( t_{.025} \) for 9 df = 2.262

\[ LSD = t_{0.025} \sqrt{MSE \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = t_{0.025} \sqrt{4.889 \left( \frac{1}{4} + \frac{1}{4} \right)} = 2.262 \sqrt{2.445} = 3.54 \]

Since \( |\bar{x}_i - \bar{x}| = |23 - 21| = 2 < 3.54 \), there does not appear to be any significant difference between the means for manufacturer 1 and manufacturer 3.
16. \( \bar{x}_1 - \bar{x}_2 \pm \text{LSD} \)

23 - 28 ± 3.54

-5 ± 3.54 = -8.54 to -1.46

18. a. \( H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \)

\( H_a: \) Not all the treatment population means are equal

Using data given in the problem, the Excel Single Factor ANOVA Output follows:

**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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>57.765</td>
<td>3</td>
<td>19.255</td>
<td>19.99481</td>
<td>0.000003</td>
</tr>
<tr>
<td>Within Groups</td>
<td>19.26</td>
<td>20</td>
<td>0.963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>77.025</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine 1</td>
<td>6</td>
<td>42.6</td>
<td>7.1</td>
<td>1.208</td>
</tr>
<tr>
<td>Machine 2</td>
<td>6</td>
<td>54.6</td>
<td>9.1</td>
<td>0.928</td>
</tr>
<tr>
<td>Machine 3</td>
<td>6</td>
<td>59.4</td>
<td>9.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Machine 4</td>
<td>6</td>
<td>68.4</td>
<td>11.4</td>
<td>1.016</td>
</tr>
</tbody>
</table>

\[ \overline{x} = (7.1 + 9.1 + 9.9 + 11.4)/4 = 9.375 \]  Note: When the sample sizes are the same, the overall s

\[ SSTR = \sum_{j=1}^{k} n_j \left( \overline{x}_j - \overline{x} \right)^2 \]

\[ = 6(7.1 - 9.375)^2 + 6(9.1 - 9.375)^2 + 6(9.9 - 9.375)^2 + 6(11.4 - 9.375)^2 = 57.765 \]

\[ MSTR = \frac{SSTR}{(k - 1)} = \frac{57.765}{3} = 19.255 \]

\[ SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 \]

\[ = 5(1.2081) + 5(.928) + 5(.70) + 5(1.016) = 19.26 \]

\[ MSE = \frac{SSE}{(n_T - k)} = \frac{19.26}{24 - 4} = .963 \]

\[ F = \frac{MSTR}{MSE} = \frac{19.26}{.963} = 19.99 \]

Using \( F \) table (3 degrees of freedom numerator and 20 denominator), \( p \)-value is less than .01

Using Excel, the \( p \)-value corresponding to \( F = 19.99 \) is \( \text{F.DIST.RT}(19.99,3,20) = .0000 \).
Because $p$-value $\leq \alpha = .05$, we reject the null hypothesis that the mean time between breakdowns is the same for the four machines.

b. $t_{0.025}$ for 20 df $= 2.086$

$$\text{LSD} = t_{\alpha/2} \sqrt{\frac{\text{MSE}}{n_i + n_j}} = 2.086 \sqrt{0.963 \left(\frac{1}{6} + \frac{1}{6}\right)} = 2.086 \sqrt{0.321} = 1.18$$

$$|\bar{x}_i - \bar{x}_j| = |9.1 - 11.4| = 2.3 > \text{LSD}; \text{ significant difference}$$

20. a. $H_0: \mu_1 = \mu_2 = \mu_3$

$H_a$: Not all the treatment population means are equal

To use Excel’s Single Factor ANOVA Tool we must first create three columns for the attendance data; one column for the attendance data for the North division, one column for the attendance data for the South division, and one column for the attendance data for the West division and then using datafile Triple-A, copy the attendance figures from the various teams into the appropriate columns. Once this is done, Excel’s Single Factor ANOVA Tool can be used to test for any significant difference in the mean attendance for the three divisions.

The Excel Single Factor ANOVA output is shown below:

<p>| SUMMARY |
|----------|----------|-----------|--------|----------|</p>
<table>
<thead>
<tr>
<th></th>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>646213</td>
<td>7702.167</td>
<td></td>
<td>1692873.8</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>422262</td>
<td>5565.5</td>
<td></td>
<td>1625453.7</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>433719</td>
<td>8429.75</td>
<td></td>
<td>324862.92</td>
<td></td>
</tr>
</tbody>
</table>

| ANOVA |
|-------|-------|--------|---------|----------|
| Source of Variation | SS     | df  | MS      | $F$     | $P$-value |
| Between Groups      | 18109727 | 2   | 9054863 | 6.9578 | 0.0111   |
| Within Groups       | 14315319 | 11  | 1301393 |         |          |
| Total               | 32425405 | 13  |         |         |          |

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

Because $p$-value $= .0111 \leq \alpha = .05$, we reject the null hypothesis that the mean attendance values are equal.

b. $n_1 = 6 \quad n_2 = 4 \quad n_3 = 4$

$t_{\alpha/2}$ is based upon 11 degrees of freedom $= 2.201$

Comparing North and South

$$\text{LSD} = t_{0.05} \sqrt{\frac{\text{MSE}}{n_i + n_j}} = 2.201 \sqrt{1.301,393 \left(\frac{1}{6} + \frac{1}{4}\right)} = 1620.76$$
Comparing North and West

\[ \text{LSD} = t_{0.05} \sqrt{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \]

\[ = 2.201 \sqrt{1,301,393 \left( \frac{1}{6} + \frac{1}{4} \right)} = 1620.76 \]

\[ |7702 - 5566| = 2136 > \text{LSD}; \text{significant difference} \]

Comparing South and West

\[ \text{LSD} = t_{0.05} \sqrt{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \]

\[ = 2.201 \sqrt{1,301,393 \left( \frac{1}{6} + \frac{1}{4} \right)} = 1775.45 \]

\[ |5566 - 8430| = 2864 > \text{LSD}; \text{significant difference} \]

The difference in the mean attendance among the three divisions is due to the low attendance in the South division.

21.  

\[ H_0: \mu_1 = \mu_2 = \mu_3 \]

\[ H_a: \text{Not all the treatment population means are equal} \]

Treatment Means (Columns):

\[ \bar{x}_1 = 13.6 \quad \bar{x}_2 = 11.0 \quad \bar{x}_3 = 10.6 \]

Block Means (Rows):

\[ \bar{x}_1 = 9 \quad \bar{x}_2 = 7.667 \quad \bar{x}_3 = 15.667 \quad \bar{x}_4 = 18.667 \quad \bar{x}_5 = 7.667 \]

Overall Mean:

\[ \bar{x} = \frac{\sum x_{ij}}{n_T} = \frac{176}{15} = 11.733 \]

Step 1

\[ \text{SST} = \sum_i \sum_j (x_{ij} - \bar{x})^2 = (10 - 11.733)^2 + (9 - 11.733)^2 + \cdots + (8 - 11.733)^2 = 354.933 \]

Step 2

\[ \text{SSTR} = b \sum_j (\bar{x}_j - \bar{x})^2 = 5 \left[ (13.6 - 11.733)^2 + (11.0 - 11.733)^2 + (10.6 - 11.733)^2 \right] = 26.533 \]
Step 3

$$SSBL = k \sum_i (\bar{x}_i - \bar{x})^2 = 3 \left[ (9 - 11.733)^2 + (7.667 - 11.733)^2 + (15.667 - 11.733)^2 + (18.667 - 11.733)^2 + (7.667 - 11.733)^2 \right] = 312.267$$

Step 4

$$SSE = SST - SSTR - SSBL = 354.933 - 26.533 - 312.267 = 16.133$$

Treatments degrees of freedom = k-1 = 3-1 = 2,
where k = the number of factors/treatments/samples being compared

Blocks df = b-1 = 5-1 = 4

Error df = (k-1)(b-1) = 2*4 = 8

Total observations = nT = k*b = 3*5 = 15

Total df = nT-1 = 15-1 = 14

$$MSTR = \frac{SSTR}{k-1} = \frac{26.533}{2} = 13.267$$

$$MSB = \frac{SSB}{b-1} = \frac{312.267}{4} = 78.067$$

$$MSE = \frac{SSE}{nT-k} = \frac{16.133}{8} = 2.017$$

$$F = \frac{MSTR}{MSE} = \frac{13.267}{2.017} = 6.58$$

Using data given in the problem, the Excel ANOVA (Two-Factor Without Replication) tool can be used to generate table (Note that in the Excel generated output, the Rows Variation is the Blocks and Columns Variation is Treatments.), or values can be filled in from calculations above

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>26.533</td>
<td>2</td>
<td>13.267</td>
<td>6.58</td>
<td>.0204</td>
</tr>
<tr>
<td>Blocks</td>
<td>312.267</td>
<td>4</td>
<td>78.067</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>16.133</td>
<td>8</td>
<td>2.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>354.933</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using $F$ table (2 degrees of freedom numerator and 8 denominator), p-value is between .01 and .025

Using Excel, the p-value corresponding to $F = 6.58 = F.DIST.RT(6.58,2,8) = .0204$.

Because $p$-value $\leq \alpha = .05$, we reject the null hypothesis that the means of the three treatments are equal.

22. Treatments $= k = 5$; Blocks $= b = 3$

$H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$H_a$: Not all the treatment population means are equal

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>310</td>
<td>4</td>
<td>77.5</td>
<td>17.71</td>
<td>.0005</td>
</tr>
<tr>
<td>Blocks</td>
<td>85</td>
<td>2</td>
<td>42.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>35</td>
<td>8</td>
<td>4.375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>430</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
SSE = SST – SSTR - SSB = 430 – 310 - 85 = 35

Treatments degrees of freedom = k-1 = 5-1 = 4,
where k = the number of factors/treatments/samples being compared

Blocks df = b-1 = 3-1 = 2

Error df = (k-1)(b-1) = 4*2 = 8

Total observations = n_T = k*b = 3*5 = 15

Total df = n_T-1 = 15-1 = 14

MSTR = SSTR/(k-1) = 310/4 = 77.5

MSB = SSB/(b-1) = 85/2 = 42.5

MSE = SSE/(n_T-k) = 35/8 = 4.375

\[ F = \frac{MSTR}{MSE} = \frac{77.5}{4.375} = 17.71 \]

Using \( F \) table (4 degrees of freedom numerator and 8 denominator), \( p \)-value is less than .01

Using Excel, the \( p \)-value corresponding to \( F = 17.71 \) = F.DIST.RT(17.71,4,8) = .0005.

Because \( p \)-value \( \leq \alpha = .05 \), we reject the null hypothesis that the means of the treatments are equal.

24. \( H_0: \mu_1 = \mu_2 \)

\( H_a: \) Not all the treatment population means are equal

\text{Treatment Means (Columns):}
\[ \bar{x}_1 = 56 \quad \bar{x}_2 = 44 \]

\text{Block Means (Rows):}
\[ \bar{x}_1 = 46 \quad \bar{x}_2 = 49.5 \quad \bar{x}_3 = 54.5 \]

\text{Overall Mean:}
\[ \bar{x}_{ij} = \frac{\sum x_{ij}}{n_T} = 300/6 = 50 \]

\text{Step 1}
\[ \text{SST} = \sum_j \sum_i (x_{ij} - \bar{x})^2 = (50 - 50)^2 + (42- 50)^2 + (55- 50)^2 + (44- 50)^2 + (63 - 50)^2 + (46 - 50)^2 = 310 \]

\text{Step 2}
\[ \text{SSTR} = b \sum_j (\bar{x}_j - \bar{x})^2 = 3 \left[ (56 - 50)^2 + (44 - 50)^2 \right] = 216 \]

\text{Step 3}
\[ \text{SSBL} = k \sum_i (\bar{x}_i - \bar{x})^2 = 2 \left[ (46 - 50)^2 + (49.5 - 50)^2 + (54.5 - 50)^2 \right] = 73 \]

\text{Step 4}
\[ \text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} = 310 - 216 - 73 = 21 \]
Treatments degrees of freedom = k-1 = 2-1 = 1,  
where k = the number of factors/treatments/samples being compared  
Blocks df = b-1 = 3-1 = 2  
Error df = (k-1)(b-1) = 1*2 = 2  
Total observations = nT = k*b = 2*3 = 6  
Total df = nT-1 = 6-1 = 5  

MSTR = SSTR/(k-1) = 216/1 = 216  
MSB = SSB/(b-1) = 73/2 = 36.5  
MSE = SSE/(nT-k) = 21/2 = 10.5  

F = MSTR/MSE = 216/10.5 = 20.57  

Using data given in the problem, the Excel ANOVA (Two-Factor Without Replication) tool can be used to generate table (Note that in the Excel generated output, the Rows Variation is the Blocks and Columns Variation is Treatments.), or values can be filled in from calculations above

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>216</td>
<td>1</td>
<td>216</td>
<td>20.57</td>
<td>.0453</td>
</tr>
<tr>
<td>Blocks</td>
<td>73</td>
<td>2</td>
<td>36.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>21</td>
<td>2</td>
<td>10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>310</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using F table (1 degree of freedom numerator and 2 denominator), p-value is between .025 and .05  
Using Excel, the p-value corresponding to F = 20.57 = F.DIST.RT(20.57,1,2) = .0453.  

Because p-value ≤ α = .05, we reject the null hypothesis that the mean tune-up times are the same for both analyzers.

26. a.  
H₀: μ₁ = μ₂ = μ₃  
H₁: Not all the treatment population means are equal

Treatment Means (Columns):  
\[ \bar{x}_1 = 502 \quad \bar{x}_2 = 515 \quad \bar{x}_3 = 494 \]

Block Means (Rows):  
\[ \bar{x}_1 = 530 \quad \bar{x}_2 = 590 \quad \bar{x}_3 = 458 \quad \bar{x}_4 = 560 \quad \bar{x}_5 = 448 \quad \bar{x}_6 = 436 \]

Overall Mean:  
\[ \bar{x} = \frac{\sum x_i}{n_T} = \frac{9066}{18} = 503.67 \]

Step 1  
\[ \text{SST} = \sum \sum (x_{ij} - \bar{x})^2 = (526 - 503.67)^2 + (534 - 503.67)^2 + \cdots + (420 - 503.67)^2 = 65,798 \]

Step 2  
\[ \text{SSTR} = b \sum (\bar{x}_j - \bar{x})^2 = 6[(502 - 503.67)^2 + (515 - 503.67)^2 + (494 - 503.67)^2] = 1348 \]
Step 3
SSBL = \[ k \sum (x_i - \overline{x})^2 = 3 \left[ (530 – 503.67)^2 + (590 - 503.67)^2 + \cdots + (436 - 503.67)^2 \right] = 63,250 \]

Step 4
SSE = SST - SSTR - SSBL = 65,798 - 1348 - 63,250 = 1200

Treatments degrees of freedom = k-1 = 3-1 = 2,
where k = the number of factors/treatments/samples being compared
Blocks df = b-1 = 6-1 = 5
Error df = (k-1)(b-1) = 2*5 = 10
Total observations = n_T = k*b = 3*6 = 18
Total df = n_T-1 = 18-1 = 17

MSTR = SSTR/(k-1) = 1348/2 = 674
MSB = SSB/(b-1) = 63250/5 = 12650
MSE = SSE/(n_T-k) = 1200/10 = 120

\[ F = \frac{MSTR}{MSE} = \frac{674}{120} = 5.62 \]

Using SATScores datafile, the Excel ANOVA (Two-Factor Without Replication) tool can be used to generate table (Note that in the Excel generated output, the Rows Variation is the Blocks and Columns Variation is Treatments.), or values can be filled in from calculations above

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>1348</td>
<td>2</td>
<td>674</td>
<td>5.62</td>
<td>.0231</td>
</tr>
<tr>
<td>Blocks</td>
<td>63,250</td>
<td>5</td>
<td>12,650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>1200</td>
<td>10</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>65,798</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using F table (2 degrees of freedom numerator and 10 denominator), p-value is between .01 and .025.
Using Excel, the p-value corresponding to \( F = 5.62 \) = F.DIST.RT(5.62,2,10) = .0231.
Using unrounded F test statistic, the p-value = .0232

Because p-value ≤ α = .05, we reject the null hypothesis that the mean scores for the three parts of the SAT are equal.

b. The mean test scores for the three sections are 502 for critical reading; 515 for mathematics; and 494 for writing. Because the writing section has the lowest average score, this section appears to give the students the most trouble.

28. \( H_0 \): Factor means are the same; no interaction effect
\( H_a \): Not all factor population means are equal or there is an interaction effect

a = number of levels of factor A = 2
b = number of levels of Factor B = 3
r = number of replications = 2

We start by calculating the sample means of each combination of Factor A and Factor B, as well as overall Factor A means and Factor B means:
### Experimental Design and Analysis of Variance

<table>
<thead>
<tr>
<th>Factor A</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{x}_{11} = 150$</td>
<td>$\bar{x}_{12} = 78$</td>
<td>$\bar{x}_{13} = 84$</td>
<td>$\bar{x}_{.1} = 104$</td>
</tr>
<tr>
<td></td>
<td>$\bar{x}_{21} = 110$</td>
<td>$\bar{x}_{22} = 116$</td>
<td>$\bar{x}_{23} = 128$</td>
<td>$\bar{x}_{.2} = 118$</td>
</tr>
</tbody>
</table>

### Step 1

\[ SST = \sum_i \sum_j \sum_k \left( x_{ijk} - \bar{x} \right)^2 = (135 - 111)^2 + (165 - 111)^2 + \cdots + (136 - 111)^2 = 9,028 \]

### Step 2

\[ SSA = br \sum_j \left( \bar{x}_j - \bar{x} \right)^2 = 3 \times (2) \left[ (104 - 111)^2 + (118 - 111)^2 \right] = 588 \]

### Step 3

\[ SSB = ar \sum_i \left( \bar{x}_i - \bar{x} \right)^2 = 2 \times (2) \left[ (130 - 111)^2 + (97 - 111)^2 + (106 - 111)^2 \right] = 2,328 \]

### Step 4

\[ SSAB = r \sum_j \sum_i \left( \bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x} \right)^2 = 2 \left[ (150 - 104 - 130 + 111)^2 + (78 - 104 - 97 + 111)^2 + \cdots + (128 - 118 - 106 + 111)^2 \right] = 4,392 \]

### Step 5

\[ SSE = SST - SSA - SSB - SSAB = 9,028 - 588 - 2,328 - 4,392 = 1,720 \]

Factor A degrees of freedom = \( a-1 = 2-1 = 1 \)

Factor B df = \( b-1 = 3-1 = 2 \)

Interaction df = \( (a-1)(b-1) = 1*2 = 2 \)

Error df = \( ab(r-1) = 2*3*1 = 6 \)

Total observations = \( n_T = a*b*r = 2*3*2 = 12 \)

Total df = \( n_T-1 = 12-1 = 11 \)

\[ MSA = SSA/(a-1) = 588/1 = 588 \]
\[ MSB = SSB/(b-1) = 2328/2 = 1164 \]
\[ MSAB = SSAB/[(a-1)(b-1)] = 4392/2 = 2196 \]
\[ MSE = SSE/[ab(r-1)] = 1720/6 = 286.67 \]

\[ F_A = MSA/MSE = 588/286.67 = 2.05 \]
\[ F_B = MSB/MSE = 1164/286.67 = 4.06 \]
\[ F_{AB} = MSAB/MSE = 2196/286.67 = 7.66 \]
Using data given in the problem, the Excel ANOVA (Two-Factor With Replication) tool can be used to generate table (Note that in the Excel generated output, the Sample Variation is Factor A, Columns Variation is Factor B, and Within is Error.), or values can be filled in from calculations above.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>588</td>
<td>1</td>
<td>588</td>
<td>2.05</td>
<td>.2022</td>
</tr>
<tr>
<td>Factor B</td>
<td>2328</td>
<td>2</td>
<td>1164</td>
<td>4.06</td>
<td>.0767</td>
</tr>
<tr>
<td>Interaction</td>
<td>4392</td>
<td>2</td>
<td>2196</td>
<td>7.66</td>
<td>.0223</td>
</tr>
<tr>
<td>Error</td>
<td>1720</td>
<td>6</td>
<td>286.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9028</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor A: $F = 2.05$

Using F table (1 degree of freedom numerator and 6 denominator), $p$-value is greater than .10
Using Excel, the $p$-value corresponding to $F = 2.05 = F.DIST.RT(2.05,1,6) = .2022$.
Using unrounded F test statistic, the p-value = .2021
Because $p$-value $> \alpha = .05$, Factor A is not significant

Factor B: $F = 4.06$

Using F table (2 degrees of freedom numerator and 6 denominator), $p$-value is between .05 and .10
Using Excel, the $p$-value corresponding to $F = 4.06 = F.DIST.RT(4.06,2,6) = .0767$.
Because $p$-value $> \alpha = .05$, Factor B is not significant

Interaction: $F = 7.66$

Using F table (2 degrees of freedom numerator and 6 denominator), $p$-value is between .01 and .025
Using Excel, the $p$-value corresponding to $F = 7.66 = F.DIST.RT(7.66,2,6) = .0223$.
Because $p$-value $\leq \alpha = .05$, Interaction is significant

30. $H_0$: Factor means are the same; no interaction effect
$H_1$: Not all factor population means are equal or there is an interaction effect

$a =$ number of levels of factor A = 3
$b =$ number of levels of Factor B = 2
$r =$ number of replications = 2

We start by calculating the sample means of each combination of Factor A and Factor B, as well as overall Factor A means and Factor B means:

Factor A is advertising design; Factor B is size of advertisement.
Experimental Design and Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>Factor B</th>
<th>Factor A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>A</td>
<td>$\bar{x}_{11} = 10$</td>
<td>$\bar{x}_{12} = 10$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>$\bar{x}_{21} = 18$</td>
<td>$\bar{x}_{22} = 28$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>$\bar{x}_{31} = 14$</td>
<td>$\bar{x}_{32} = 16$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor B</td>
<td>Means</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\bar{x}_B = 14$</td>
<td>$\bar{x}_C = 18$</td>
</tr>
</tbody>
</table>

Step 1

$$SST = \sum_{i} \sum_{j} \sum_{k} (x_{ijk} - \bar{x})^2 = (8 - 16)^2 + (12 - 16)^2 + (12 - 16)^2 + \cdots + (14 - 16)^2 = 544$$

Step 2

$$SSA = b r \sum_j (\bar{x}_j - \bar{x})^2 = 2 \left( (10 - 16)^2 + (23 - 16)^2 + (15 - 16)^2 \right) = 344$$

Step 3

$$SSB = a r \sum_j (\bar{x}_j - \bar{x})^2 = 3 \left( (14 - 16)^2 + (18 - 16)^2 \right) = 48$$

Step 4

$$SSAB = r \sum_j \sum_k (\bar{x}_{jk} - \bar{x}_j - \bar{x}_k + \bar{x})^2 = 2 \left( (10 - 10 - 14 + 16)^2 + \cdots + (16 - 15 - 18 +16)^2 \right) = 56$$

Step 5

$$SSE = SST - SSA - SSB - SSAB = 544 - 344 - 48 - 56 = 96$$

Factor A degrees of freedom = a-1 = 3-1 = 2
Factor B df = b-1 = 2-1 = 1
Interaction df = (a-1)(b-1) = 2*1 = 2
Error df = ab(r-1) = 3*2*1 = 6

Total observations = nT = a*b*r = 3*2*2 = 12
Total df = nT-1 = 12-1 = 11

$$MSA = SSA/(a-1) = 344/2 = 172$$
$$MSB = SSB/(b-1) = 48/1 = 48$$
$$MSAB = SSAB/[(a-1)(b-1)] = 56/2 = 28$$
$$MSE = SSE/[ab(r-1)] = 96/6 = 16$$

$$F_A = MSA/MSE = 172/16 = 10.75$$
$$F_B = MSB/MSE = 48/16 = 3.00$$
Using data given in the problem, the Excel ANOVA (Two-Factor With Replication) tool can be used to generate table (Note that in the Excel generated output, the Sample Variation is Factor A, Columns Variation is Factor B, and Within is Error.), or values can be filled in from calculations above.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>344</td>
<td>2</td>
<td>172</td>
<td>10.75</td>
<td>.0104</td>
</tr>
<tr>
<td>Factor B</td>
<td>48</td>
<td>1</td>
<td>48</td>
<td>3.00</td>
<td>.1340</td>
</tr>
<tr>
<td>Interaction</td>
<td>56</td>
<td>2</td>
<td>28</td>
<td>1.75</td>
<td>.2519</td>
</tr>
<tr>
<td>Error</td>
<td>96</td>
<td>6</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>544</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor A: $F = 10.75$

Using $F$ table for Factor A (2 degrees of freedom numerator and 6 denominator), $p$-value is between .01 and .025

Using Excel, the $p$-value corresponding to $F = 10.75 = F.DIST.RT(10.75,2,6)$ =.0104.

Because $p$-value $\leq \alpha = .05$, Factor A is significant; there is a difference due to the type of advertisement design.

Factor B: $F = 3.00$

Using $F$ table for Factor B (1 degree of freedom numerator and 6 denominator), $p$-value is greater than .10

Using Excel, the $p$-value corresponding to $F = 3.00 = F.DIST.RT(3,1,6)$ =.1340.

Because $p$-value $> \alpha = .05$, Factor B is not significant; there is not a significant difference due to size of advertisement.

Interaction: $F = 1.75$

Using $F$ table for Interaction (2 degrees of freedom numerator and 6 denominator), $p$-value is greater than .10

Using Excel, the $p$-value corresponding to $F = 1.75 = F.DIST.RT(1.75,2,6)$ =.2519.

Because $p$-value $> \alpha = .05$, Interaction is not significant.

32. $H_0$: Factor means are the same; no interaction effect
$H_a$: Not all factor population means are equal or there is an interaction effect

a = number of levels of factor A = 4
b = number of levels of Factor B = 2
r = number of replications = 2

Using HybridTest datafile with Factor A as Class of vehicle tested (small car, midsize car, small SUV, and midsize SUV) and Factor B as Type (hybrid or conventional), the data in tabular format is created as follows.

<table>
<thead>
<tr>
<th></th>
<th>Hybrid</th>
<th>Conventional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Car</td>
<td>37</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>32</td>
</tr>
</tbody>
</table>
Next we calculate the sample means of each combination of Factor A and Factor B, as well as overall Factor A means and Factor B means:

Summary statistics for the above data are shown below:

\[
\begin{array}{cccc}
\text{Small Car} & \bar{x}_{11} = 11 & \bar{x}_{12} = 12 & \bar{x}_{1} = 11.5 \\
\text{Midsize Car} & \bar{x}_{21} = 21 & \bar{x}_{22} = 22 & \bar{x}_{2} = 21.5 \\
\text{Small SUV} & \bar{x}_{31} = 31 & \bar{x}_{32} = 32 & \bar{x}_{3} = 31.5 \\
\text{Midsize SUV} & \bar{x}_{41} = 41 & \bar{x}_{42} = 42 & \bar{x}_{4} = 41.5 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{Hybrid} & \bar{x}_1 = 30.25 & \bar{x}_2 = 23.5 & \bar{x} = 26.875 \\
\text{Conventional} & \bar{x}_1 = 40.5 & \bar{x}_2 = 30.0 & \bar{x}_1 = 35.25 \\
\end{array}
\]

Step 1
\[
SST = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x})^2 = (37 - 26.875)^2 + (44 - 26.875)^2 + \cdots + (18 - 26.875)^2 = 691.75
\]

Step 2
\[
SSA = b \sum_i (\bar{x}_i - \bar{x})^2 = 2(2) [(35.25 - 26.875)^2 + (26.75 - 26.875)^2 + (24.5 - 26.875)^2 + (21.0 - 26.875)^2] = 441.25
\]

Step 3
\[
SSB = a \sum_j (\bar{x}_j - \bar{x})^2 = 4(2) [(30.25 - 26.875)^2 + (23.5 - 26.875)^2] = 182.25
\]

Step 4
\[
SSAB = r \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = 2[(40.5 - 35.25 - 30.25 + 26.875)^2 + (30 - 35.25 - 23.5 + 26.875)^2 + \cdots + (18.5 - 21.0 - 23.5 + 26.875)^2] = 19.25
\]

Step 5
\[
SSE = SST - SSA - SSB - SSAB = 691.75 - 441.25 - 182.25 - 19.25 = 49
\]

Factor A degrees of freedom = a-1 = 4-1 = 3
Factor B df = b-1 = 2-1 = 1
Interaction df = (a-1)(b-1) = 3*1 = 3
Error df = ab(r-1) = 4*2*1 = 8

Total observations = n_T = a*b*r = 4*2*2 = 16
Total df = n_T-1 = 16-1 = 15

MSA = SSA/(a-1) = 441.25/3 = 147.083
MSB = SSB/(b-1) = 182.25/1 = 182.25
MSAB = SSAB/[(a-1)(b-1)] = 19.25/3 = 6.4167  
MSE = SSE/[ab(r-1)] = 49/8 = 6.125

\[ F_A = \frac{MSA}{MSE} = 147.083/6.125 = 24.01 \]

\[ F_B = \frac{MSB}{MSE} = 182.25/6.125 = 29.76 \]

\[ F_{AB} = \frac{MSAB}{MSE} = 6.4167/6.125 = 1.0476 \]

Using HybridTest datafile, the Excel ANOVA (Two-Factor With Replication) tool can be used to generate table (Note that in the Excel generated output, the Sample Variation is Factor A, Columns Variation is Factor B, and Within is Error.), or values can be filled in from calculations above

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>441.25</td>
<td>3</td>
<td>147.083</td>
<td>24.01</td>
<td>.0002</td>
</tr>
<tr>
<td>Factor B</td>
<td>182.25</td>
<td>1</td>
<td>182.250</td>
<td>29.76</td>
<td>.0006</td>
</tr>
<tr>
<td>Interaction</td>
<td>19.25</td>
<td>3</td>
<td>6.4167</td>
<td>1.0476</td>
<td>.4229</td>
</tr>
<tr>
<td>Error</td>
<td>49.00</td>
<td>8</td>
<td>6.125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>691.75</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor A: \( F = 24.01 \)

Using \( F \) table for Factor A (3 degree of freedom numerator and 8 denominator), p-value is less than .01

Using Excel, the p-value corresponding to \( F = 24.01 \) = F.DIST.RT(24.01,3,8) = .0002.

Factor A: Because \( p \)-value = .0002 \( \leq \alpha = .05 \), Factor A (Class) is significant

Factor B: \( F = 29.76 \)

Using \( F \) table for Factor B (1 degree of freedom numerator and 8 denominator), p-value is less than .01

Using Excel, the p-value corresponding to \( F = 29.76 \) = F.DIST.RT(29.76,1,8) = .0006.

Factor B: Because \( p \)-value = .0006 \( \leq \alpha = .05 \), Factor B (Type) is significant

Interaction: \( F = 1.0476 \)

Using \( F \) table for Interaction (3 degrees of freedom numerator and 8 denominator), p-value is greater than .10

Using Excel, the p-value corresponding to \( F = 1.0476 \) = F.DIST.RT(1.0476,3,8) = .4229.

Interaction: Because \( p \)-value = .4229 \( > \alpha = .05 \), Interaction is not significant

The class of vehicles has a significant effect on miles per gallon with cars showing more miles per gallon than SUVs. The type of vehicle also has a significant effect with hybrids having more miles per gallon than conventional vehicles. There is no evidence of a significant interaction effect.

34. \( H_0: \mu_1 = \mu_2 = \mu_3 \)

\( H_a: \) Not all the treatment population means are equal

Using data given in problem, the Excel Single Factor ANOVA Output follows:
Anova: Single Factor

### SUMMARY

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>4</td>
<td>368</td>
<td>92</td>
<td>30</td>
</tr>
<tr>
<td>y</td>
<td>4</td>
<td>388</td>
<td>97</td>
<td>6</td>
</tr>
<tr>
<td>z</td>
<td>4</td>
<td>336</td>
<td>84</td>
<td>35.33333</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>344</td>
<td>2</td>
<td>172</td>
<td>7.233645</td>
<td>0.013397</td>
</tr>
<tr>
<td>Within Groups</td>
<td>214</td>
<td>9</td>
<td>23.77778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>558</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

\[
\bar{x} = \frac{(92 + 97 + 84)}{3} = 91 \text{ Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.}
\]

\[
SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 4(92 - 91)^2 + 4(97 - 91)^2 + 4(84 - 91)^2 = 344
\]

\[
MSTR = \frac{SSTR}{(k - 1)} = \frac{344}{2} = 172
\]

\[
SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 3(30) + 3(6) + 3(35.33) = 214
\]

\[
MSE = \frac{SSE}{(n_T - k)} = \frac{214}{(12 - 3)} = 23.78
\]

\[
F = \frac{MSTR}{MSE} = \frac{172}{23.78} = 7.23
\]

Using \( F \) table (2 degrees of freedom numerator and 9 denominator), \( p \)-value is between .01 and .025

Using Excel, the \( p \)-value corresponding to \( F = 7.23 \) = F.DIST.RT(7.23,2,9) = .0134.

Because \( p \)-value \( \leq \alpha = .05 \), we reject the null hypothesis that the mean absorbency ratings for the three brands are equal.

36. \( H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \)

\( H_a: \) Not all the treatment population means are equal

Using OzoneLevels datafile, the Excel ANOVA (Two-Factor Without Replication) output is shown below:

### 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>903.025</td>
<td>9</td>
<td>100.3361</td>
<td>4.5479</td>
<td>0.0010</td>
<td>2.2501</td>
</tr>
<tr>
<td>Columns</td>
<td>160.075</td>
<td>3</td>
<td>53.3583</td>
<td>2.4186</td>
<td>0.0880</td>
<td>2.9604</td>
</tr>
</tbody>
</table>
The label Rows corresponds to the blocks in the problem (Date), and the label column corresponds to the treatments (City).

Because the $p$-value corresponding to Columns (treatments) is .0880 > $\alpha = .05$, there is no significant difference in the mean ozone level among the four cites. But, if the level of significance was $\alpha = .10$, the difference would have been significant.

38. $H_0$: $\mu_1 = \mu_2 = \mu_3$
$H_a$: Not all the treatment population means are equal

Using Assembly datafile, the Excel Single Factor ANOVA Output follows:

### SUMMARY

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>900</td>
<td>90</td>
<td>98</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>840</td>
<td>84</td>
<td>168.4444</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>810</td>
<td>81</td>
<td>159.7778</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>420</td>
<td>2</td>
<td>210</td>
<td>1.478102</td>
<td>0.245946</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3836</td>
<td>27</td>
<td>142.0741</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4256</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

<table>
<thead>
<tr>
<th></th>
<th>Method A</th>
<th>Method B</th>
<th>Method C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean</td>
<td>90</td>
<td>84</td>
<td>81</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>98.00</td>
<td>168.44</td>
<td>159.78</td>
</tr>
</tbody>
</table>

$\bar{x} = (90 + 84 + 81)/3 = 85$ Note: When the sample sizes are the same, the overall sample mean is an average of the individual sample means.

$$SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 10(90 - 85)^2 + 10(84 - 85)^2 + 10(81 - 85)^2 = 420$$

$$MSTR = SSTR / (k - 1) = 420 / 2 = 210$$

$$SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 9(98.00) + 9(168.44) + 9(159.78) = 3,836$$

$$MSE = SSE / (n_T - k) = 3,836 / (30 - 3) = 142.07$$

$$F = MSTR / MSE = 210 / 142.07 = 1.48$$
Using $F$ table (2 degrees of freedom numerator and 27 denominator), $p$-value is greater than .10
Using Excel, the $p$-value corresponding to $F = 1.48 = \text{F.DIST.RT}(1.48,2,27) = .2455$.
Using unrounded $F$ test statistic, the $p$-value = .2459

Because $p$-value > $\alpha = .05$, we cannot reject the null hypothesis that the means are equal.

40. a. $H_0$: $\mu_1 = \mu_2 = \mu_3$
$H_a$: Not all the treatment population means are equal

Treatment Means (Columns):
$\bar{x}_1 = 22.8$  $\bar{x}_2 = 24.8$  $\bar{x}_3 = 25.80$

Block Means (Rows):
$\bar{x}_1 = 19.667$  $\bar{x}_2 = 25.667$  $\bar{x}_3 = 31$  $\bar{x}_4 = 23.667$  $\bar{x}_5 = 22.333$

Overall Mean:
$\bar{x} = \frac{\sum x_{ij}}{n_T} = \frac{367}{15} = 24.467$

Step 1
$\text{SST} = \sum \sum (x_{ij} - \bar{x})^2 = (18 - 24.467)^2 + (21 - 24.467)^2 + \cdots + (24 - 24.467)^2 = 253.733$

Step 2
$\text{SSTR} = b \sum (\bar{x}_j - \bar{x})^2 = 5 \left[ (22.8 - 24.467)^2 + (24.8 - 24.467)^2 + (25.8 - 24.467)^2 \right] = 23.333$

Step 3
$\text{SSBL} = k \sum (\bar{x}_i - \bar{x})^2 = 3 \left[ (19.667 - 24.467)^2 + (25.667 - 24.467)^2 + (31 - 24.467)^2 + (23.667 - 24.467)^2 + (22.333 - 24.467)^2 \right] = 217.067$

Step 4
$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} = 253.733 - 23.333 - 217.067 = 13.333$

Treatments degrees of freedom = $k-1 = 3-1 = 2$

where $k$ = the number of factors/treatments/samples being compared
Blocks df = $b-1 = 5-1 = 4$
Error df = $(k-1)(b-1) = 2*4 = 8$

Total observations = $n_T = k*b = 3*5 = 15$
Total df = $n_T-1 = 15-1 = 14$

$\text{MSTR} = \frac{\text{SSTR}}{(k-1)} = \frac{23.333}{2} = 11.667$
$\text{MSB} = \frac{\text{SSBL}}{(b-1)} = \frac{217.067}{4} = 54.267$
$\text{MSE} = \frac{\text{SSE}}{(n_T-k)} = \frac{13.333}{8} = 1.667$

$F = \frac{\text{MSTR}}{\text{MSE}} = \frac{11.667}{1.667} = 7.00$

Using data given in the problem, the Excel ANOVA (Two-Factor Without Replication) tool can be used to generate table (Note that in the Excel generated output, the Rows Variation is the Blocks and Columns Variation is Treatments.), or values can be filled in from calculations above.
Chapter 13

### Summary

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>23.333</td>
<td>2</td>
<td>11.667</td>
<td>7.00</td>
<td>.0175</td>
</tr>
<tr>
<td>Blocks</td>
<td>217.067</td>
<td>4</td>
<td>54.267</td>
<td>32.56</td>
<td>.0175</td>
</tr>
<tr>
<td>Error</td>
<td>13.333</td>
<td>8</td>
<td>1.667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>253.733</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using $F$ table (2 degrees of freedom numerator and 8 denominator), $p$-value is between .01 and .025

Using Excel, the $p$-value corresponding to $F = 7.00= F.DIST.RT(7,2,8) = .0175$.

Because $p$-value $\leq \alpha = .05$, we reject the null hypothesis that the mean miles per gallon ratings for the three brands of gasoline are equal.

**b.**

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_a$: Not all the treatment population means are equal

Completely Randomized Design is Single Factor ANOVA (no blocks)

Using data given in the problem, the Excel Single Factor ANOVA Output follows:

**Anova: Single Factor**

### SUMMARY

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>5</td>
<td>114</td>
<td>22.8</td>
<td>21.2</td>
</tr>
<tr>
<td>II</td>
<td>5</td>
<td>124</td>
<td>24.8</td>
<td>9.2</td>
</tr>
<tr>
<td>III</td>
<td>5</td>
<td>129</td>
<td>25.8</td>
<td>27.2</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>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>23.333</td>
<td>2</td>
<td>11.667</td>
<td>0.607639</td>
<td>0.56057</td>
</tr>
<tr>
<td>Within Groups</td>
<td>230.4</td>
<td>12</td>
<td>19.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>253.733</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that Between Groups Variation is the Treatments and Within Groups Variation is Error.

<table>
<thead>
<tr>
<th>Groups</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean</td>
<td>22.8</td>
<td>24.8</td>
<td>25.8</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>21.2</td>
<td>9.2</td>
<td>27.2</td>
</tr>
</tbody>
</table>

$\bar{x} = (22.8 + 24.8 + 25.8) / 3 = 24.467$

**Note:** When the sample sizes are the same, the overall sample mean is an average of the individual sample means.

\[
SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 5(22.8 - 24.467)^2 + 5(24.8 - 24.467)^2 + 5(25.8 - 24.467)^2 = 23.333
\]

\[
MSTR = SSTR / (k - 1) = 23.333 / 2 = 11.667
\]

\[
SSE = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 4(21.2) + 4(9.2) + 4(27.2) = 230.4
\]
MSE = SSE / (n_T - k) = 230.4 / (15 - 3) = 19.2

F = MSTR / MSE = 11.667 / 19.2 = .61

Using F table (2 degrees of freedom numerator and 12 denominator), p-value is greater than .10

Using Excel, the p-value corresponding to \( F = .61 \) = F.DIST.RT(.61,2,12) = .5594.

Using unrounded F test statistic, the pvalue = .5606

Because \( p \)-value > \( \alpha = .05 \), we cannot reject the null hypothesis that the mean miles per gallon ratings for the three brands of gasoline are equal.

Thus, we must remove the block effect in order to detect a significant difference due to the brand of gasoline. The following table illustrates the relationship between the randomized block design and the completely randomized design.

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>Randomized Block Design</th>
<th>Completely Randomized Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>SST</td>
<td>253.733</td>
<td>253.733</td>
</tr>
<tr>
<td>SSTR</td>
<td>23.333</td>
<td>23.333</td>
</tr>
<tr>
<td>SSBL</td>
<td>217.067</td>
<td>does not exist</td>
</tr>
<tr>
<td>SSE</td>
<td>13.333</td>
<td>230.4</td>
</tr>
</tbody>
</table>

Note that SSE for the completely randomized design is the sum of SSBL (217.02) and SSE (13.38) for the randomized block design. This illustrates that the effect of blocking is to remove the block effect from the error sum of squares; thus, the estimate of \( \sigma^2 \) for the randomized block design is substantially smaller than it is for the completely randomized design.

42. \( H_0: \mu_1 = \mu_2 = \mu_3 \)
    \( H_a: \) Not all the treatment population means are equal

Using HoustonAstros datafile, the Excel ANOVA (Two-Factor Without Replication) output is shown below:

<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>79329936</td>
<td>6</td>
<td>13221656</td>
<td>0.7002</td>
<td>0.6552</td>
<td>2.9961</td>
</tr>
<tr>
<td>Columns</td>
<td>1.01E+08</td>
<td>2</td>
<td>50734841</td>
<td>2.6870</td>
<td>0.1086</td>
<td>3.8853</td>
</tr>
<tr>
<td>Error</td>
<td>2.27E+08</td>
<td>12</td>
<td>18881427</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4.07E+08</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The label Rows corresponds to the blocks in the problem (Opponent), and the label column corresponds to the treatments (Day).

Because the \( p \)-value corresponding to Columns (treatments) is \( .1086 > \alpha = .05 \), there is no significant difference in the mean attendance per game for games played on Friday, Saturday, and Sunday. These data do not suggest a particular day on which the Astros should schedule these promotions.

44. \( H_0: \) Factor means are the same; no interaction effect
    \( H_a: \) Not all factor population means are equal or there is an interaction effect
Chapter 13

\[ a = \text{number of levels of factor A} = 2 \]
\[ b = \text{number of levels of Factor B} = 2 \]
\[ r = \text{number of replications} = 2 \]

We start by calculating the sample means of each combination of Factor A and Factor B, as well as overall Factor A means and Factor B means:

<table>
<thead>
<tr>
<th>Factor A</th>
<th>Factor B Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine 1</td>
<td>( \bar{x}<em>{11} = 32 ) ( \bar{x}</em>{12} = 28 ) ( \bar{x}_{.,1} = 30 )</td>
</tr>
<tr>
<td>Machine 2</td>
<td>( \bar{x}<em>{21} = 21 ) ( \bar{x}</em>{22} = 26 ) ( \bar{x}_{.,2} = 23.5 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor B Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
</tr>
<tr>
<td>Automatic</td>
</tr>
<tr>
<td>( \bar{x}_1 = 26.5 )</td>
</tr>
<tr>
<td>( \bar{x}_2 = 27 )</td>
</tr>
<tr>
<td>( \bar{x} = 26.75 )</td>
</tr>
</tbody>
</table>

**Step 1**

\[ \text{SST} = \sum_{i} \sum_{j} \sum_{k} (x_{ijk} - \bar{x})^2 = (30 - 26.75)^2 + (34 - 26.75)^2 + \ldots + (28 - 26.75)^2 = 151.5 \]

**Step 2**

\[ \text{SSA} = br \sum_{i} (\bar{x}_i - \bar{x})^2 = 2 (\sum 26.75)^2 + (23.5 - 26.75)^2 ] = 84.5 \]

**Step 3**

\[ \text{SSB} = ar \sum_{j} (\bar{x}_j - \bar{x})^2 = 2 (\sum (26.5 - 26.75)^2 + (27 - 26.75)^2 ] = 0.5 \]

**Step 4**

\[ \text{SSAB} = r \sum_{i} \sum_{j} (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = 2[(32 - 30 - 26.5 + 26.75)^2 + \ldots + (26 - 23.5 - 27 + 26.75)^2] \]
\[ = 40.5 \]

**Step 5**

\[ \text{SSE} = \text{SST} - \text{SSA} - \text{SSB} - \text{SSAB} = 151.5 - 84.5 - 0.5 - 40.5 = 26 \]

Factor A degrees of freedom = \( a-1 = 2-1 = 1 \)
Factor B df = \( b-1 = 2-1 = 1 \)
Interaction df = \( (a-1)(b-1) = 1*1 = 1 \)
Error df = \( ab(r-1) = 2*2*1 = 4 \)
Total observations = \( n_T = a*b*r = 2*2*2 = 8 \)
Total df = \( n_T-1 = 8-1 = 7 \)

\[ \text{MSA} = \frac{\text{SSA}}{a-1} = \frac{84.5}{1} = 84.5 \]
\[ \text{MSB} = \frac{\text{SSB}}{b-1} = \frac{.5}{1} = .5 \]
MSAB = SSAB/[(a-1)(b-1)] = 40.5/1 = 40.5
MSE = SSE/[ab(r-1)] = 26/4 = 6.5

\[ F_A = \frac{MSA}{MSE} = \frac{84.5}{6.5} = 13 \]
\[ F_B = \frac{MSB}{MSE} = \frac{.5}{6.5} = .0769 \]
\[ F_{AB} = \frac{MSAB}{MSE} = \frac{40.5}{6.5} = 6.231 \]

Using data given in the problem, the Excel ANOVA (Two-Factor With Replication) tool can be used to generate table (Note that in the Excel generated output, the Sample Variation is Factor A, Columns Variation is Factor B, and Within is Error.), or values can be filled in from calculations above.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>84.5</td>
<td>1</td>
<td>84.5</td>
<td>13</td>
<td>.0226</td>
</tr>
<tr>
<td>Factor B</td>
<td>.5</td>
<td>1</td>
<td>.5</td>
<td>.0769</td>
<td>.7953</td>
</tr>
<tr>
<td>Interaction</td>
<td>40.5</td>
<td>1</td>
<td>40.5</td>
<td>6.231</td>
<td>.0670</td>
</tr>
<tr>
<td>Error</td>
<td>26</td>
<td>4</td>
<td>6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>151.5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor A: \( F = 13 \)

Using \( F \) table for Factor A (1 degrees of freedom numerator and 4 denominator), \( p \)-value is between .01 and .025

Using Excel, the \( p \)-value corresponding to \( F = 13 = F.DIST.RT(13,1,4) = .0226. \)

Because \( p \)-value \( \leq \alpha = .05 \), Factor A (machine) is significant.

Factor B: \( F = .0769 \)

Using \( F \) table for Factor B (1 degrees of freedom numerator and 4 denominator), \( p \)-value is greater than .10

Using Excel, the \( p \)-value corresponding to \( F = .0769 = F.DIST.RT(.0769,1,4) = .7953. \)

Because \( p \)-value > \( \alpha = .05 \), Factor B (loading system) is not significant.

Interaction: \( F = 6.231 \)

Using \( F \) table for Interaction (1 degrees of freedom numerator and 4 denominator), \( p \)-value is between .05 and .10

Using Excel, the \( p \)-value corresponding to \( F = 6.231 = F.DIST.RT(6.231,1,4) = .0670. \)

Because \( p \)-value > \( \alpha = .05 \), Interaction is not significant.
Chapter 14
Simple Linear Regression

Solutions:

1. a.

b. There appears to be a positive linear relationship between $x$ and $y$.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between $x$ and $y$; in part (d) we will determine the equation of a straight line that “best” represents the relationship according to the least squares criterion.

d. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{15}{5} = 3 \quad \bar{y} = \frac{\Sigma y_i}{n} = \frac{40}{5} = 8$

$\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 26 \quad \Sigma (x_i - \bar{x})^2 = 10$

$b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{26}{10} = 2.6$

$b_0 = \bar{y} - b_1\bar{x} = 8 - (2.6)(3) = 0.2$

$\hat{y} = 0.2 + 2.6x$

e. $\hat{y} = 0.2 + 2.6(4) = 10.6$
2. a. 

b. There appears to be a negative linear relationship between $x$ and $y$.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between $x$ and $y$; in part (d) we will determine the equation of a straight line that “best” represents the relationship according to the least squares criterion.

d. 

\[
\bar{x} = \frac{\sum x_i}{n} = \frac{55}{5} = 11 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{175}{5} = 35
\]

\[
\Sigma (x_i - \bar{x})(y_i - \bar{y}) = -540 \quad \Sigma (x_i - \bar{x})^2 = 180
\]

\[
b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = -\frac{540}{180} = -3
\]

\[
b_0 = \bar{y} - b_1 \bar{x} = 35 - (-3)(11) = 68
\]

\[
\hat{y} = 68 - 3x
\]

e. \[\hat{y} = 68 - 3(10) = 38\]
4. a.

b. There appears to be a positive linear relationship between the percentage of women working in the
five companies (x) and the percentage of management jobs held by women in that company (y)

c. Many different straight lines can be drawn to provide a linear approximation of the relationship
between x and y; in part (d) we will determine the equation of a straight line that “best” represents
the relationship according to the least squares criterion.

d. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{300}{5} = 60 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{215}{5} = 43 \]

\[ \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 624 \quad \Sigma (x_i - \bar{x})^2 = 480 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{624}{480} = 1.3 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 43 - 1.3(60) = -35 \]

\[ \hat{y} = -35 + 1.3x \]

e. \[ \hat{y} = -35 + 1.3x = -35 + 1.3(60) = 43% \]
6. a. The scatter diagram indicates a positive linear relationship between \( x \) = average number of passing yards per attempt and \( y \) = the percentage of games won by the team.

6. b. 

\[
\bar{x} = \frac{\sum x_i}{n} = 68/10 = 6.8 \quad \bar{y} = \frac{\sum y_i}{n} = 464/10 = 46.4
\]

\[
\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 121.6 \quad \Sigma(x_i - \bar{x})^2 = 7.08
\]

\[
b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{121.6}{7.08} = 17.1751
\]

\[
b_0 = \bar{y} - b_1\bar{x} = 46.4 - (17.1751)(6.8) = -70.391
\]

\[
y = -70.391 + 17.1751x
\]

d. The slope of the estimated regression line is approximately 17.2. So, for every increase of one yard in the average number of passes per attempt, the percentage of games won by the team increases by 17.2%.

e. With an average number of passing yards per attempt of 6.2, the predicted percentage of games won is \( \hat{y} = -70.391 + 17.175(6.2) = 36\% \). With a record of 7 wins and 9 loses, the percentage of wins that the Kansas City Chiefs won is 43.8 or approximately 44\%. Considering the small data size, the prediction made using the estimated regression equation is not too bad.
8. a.

b. The scatter diagram indicates a positive linear relationship between \( x = \) speed of execution rating and \( y = \) overall satisfaction rating for electronic trades.

c. \[
\bar{x} = \frac{\Sigma x_i}{n} = 36.3/11 = 3.3 \quad \bar{y} = \frac{\Sigma y_i}{n} = 35.2/11 = 3.2
\]

\[
\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 2.36 \quad \Sigma x_i - \bar{x})^2 = 2.6
\]

\[
b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{2.36}{2.6} = 0.9077
\]

\[
b_0 = \bar{y} - b_1\bar{x} = 3.2 - (0.9077)(3.3) = 0.2046
\]

\[
y = 0.2046 + 0.9077x
\]

d. The slope of the estimated regression line is approximately 0.9077. So, a one unit increase in the speed of execution rating will increase the overall satisfaction rating by approximately 0.9 points.

e. The average speed of execution rating for the other brokerage firms is 3.38. Using this as the new value of \( x \) for Zecco.com, we can use the estimated regression equation developed in part (c) to estimate the overall satisfaction rating corresponding to \( x = 3.38 \).

\[
y = 0.2046 + 0.9077x = 0.2046 + 0.9077(3.38) = 3.27
\]

Thus, an estimate of the overall satisfaction rating when \( x = 3.38 \) is approximately 3.3.
The scatter diagram indicates a positive linear relationship between $x = \text{age of wine}$ and $y = \text{price of a 750 ml bottle of wine}$. In other words, the price of the wine increases with age.

\[
\bar{x} = \frac{\sum x_i}{n} = \frac{317}{10} = 31.7 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{2294}{10} = 229.4
\]

\[
\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 3838.2 \quad \Sigma(x_i - \bar{x})^2 = 552.1
\]

\[
b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{3838.2}{552.1} = 6.952
\]

\[
b_0 = \bar{y} - b_1 \bar{x} = 229.4 - (6.952)(31.7) = 9.0216
\]

\[
\hat{y} = b_0 + b_1 x = 9.0216 + 6.952 x
\]

The slope of the estimated regression line is approximately 6.95. So, for every additional year of age, the price of the wine increases by $6.95.
12. a.

b. The scatter diagram indicates a somewhat positive linear relationship between \( x \) = percentage return of the S&P 500 and \( y \) = percentage return for Coca-Cola.

c. \[
\bar{x} = \frac{\sum x_i}{n} = \frac{2.00}{10} = 0.2 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{15}{10} = 1.5
\]
\[
\sum (x_i - \bar{x})(y_i - \bar{y}) = 95 \quad \sum (x_i - \bar{x})^2 = 179.6
\]
\[b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{95}{179.6} = 0.5290
\]
\[b_0 = \bar{y} - b_1 \bar{x} = 1.5 - (0.5290)(0.2) = 1.3942
\]
\[\hat{y} = 1.3942 + 0.529x
\]

d. A one percent increase in the percentage return of the S&P 500 will result in a 0.529 increase in the percentage return for Coca-Cola.

e. The beta of 0.529 for Coca-Cola differs somewhat from the beta of 0.82 reported by Yahoo Finance. This is likely due to differences in the period over which the data were collected and the amount of data used to calculate the beta. Note: Yahoo uses the last five years of monthly returns to calculate a beta.
14. a.

b. The scatter diagram indicates a positive linear relationship between \( x = \text{price (}) \) and \( y = \text{overall rating}. \)

c. \( \bar{x} = \frac{\sum x_i}{n} = \frac{4660}{20} = 233 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{1400}{20} = 70 \)

\[ \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 8100 \quad \Sigma (x_i - \bar{x})^2 = 127,420 \]

\[ b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{8100}{127,420} = 0.06357 \]

\[ b_0 = \bar{y} - b_1 \bar{x} = 70 - (0.06357)(233) = 55.188 \]

\[ \hat{y} = 55.188 + 0.06357x \]

d. We can use the estimated regression equation developed in part (c) to estimate the overall satisfaction rating corresponding to \( x = 200. \)

\[ \hat{y} = 55.188 + 0.06357x = 55.188 + 0.06357(200) = 67.9 \]

Thus, an estimate of the overall rating when \( x = 200 \) is approximately 68.

15. a. The estimated regression equation and the mean for the dependent variable are:

\[ \hat{y}_i = 0.2 + 2.6x_i \quad \bar{y} = 8 \]

The sum of squares due to error and the total sum of squares are

\[ \text{SSE} = \Sigma (y_i - \hat{y}_i)^2 = 12.40 \quad \text{SST} = \Sigma (y_i - \bar{y})^2 = 80 \]

Thus, \( \text{SSR} = \text{SST} - \text{SSE} = 80 - 12.4 = 67.6 \)
b. \( r^2 = \frac{SSR}{SST} = \frac{67.6}{80} = .845 \)

The least squares line provided a very good fit; 84.5% of the variability in \( y \) has been explained by the least squares line.

c. \( r_{xy} = \sqrt{.845} = +.9192 \)

16. a. The estimated regression equation and the mean for the dependent variable are:

\[ \hat{y}_i = 68 - 3x \quad \bar{y} = 35 \]

The sum of squares due to error and the total sum of squares are

\[ SSE = \sum (y_i - \hat{y}_i)^2 = 230 \quad SST = \sum (y_i - \bar{y})^2 = 1850 \]

Thus, SSR = SST - SSE = 1850 - 230 = 1620

b. \( r^2 = \frac{SSR}{SST} = \frac{1620}{1850} = .876 \)

The least squares line provided an excellent fit; 87.6% of the variability in \( y \) has been explained by the estimated regression equation.

c. \( r_{xy} = \sqrt{.876} = -.936 \)

Note: the sign for \( r \) is negative because the slope of the estimated regression equation is negative. \( (b_1 = -3) \)

18. a. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{600}{6} = 100 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{330}{6} = 55 \]

\[ SST = \sum (y_i - \bar{y})^2 = 1800 \quad SSE = \sum (y_i - \hat{y}_i)^2 = 287.624 \]

\[ SSR = SST - SSE = 1800 - 287.624 = 1512.376 \]

b. \[ r^2 = \frac{SSR}{SST} = \frac{1512.376}{1800} = .84 \]

The least squares line provided a very good fit; 84% of the variability in \( y \) has been explained by the least squares line.

c. \( r = \sqrt{r^2} = \sqrt{.84} = .917 \)

20. a. \[ \bar{x} = \frac{\sum x_i}{n} = \frac{160}{10} = 16 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{55,500}{10} = 5550 \]

\[ \Sigma(x_i - \bar{x})(y_i - \bar{y}) = -31,284 \quad \Sigma(x_i - \bar{x})^2 = 21.74 \]

\[ b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-31,284}{21.74} = -1439 \]

\[ b_0 = \bar{y} - b_1\bar{x} = 5550 - (-1439)(16) = 28,574 \]

\[ \hat{y} = 28,574 - 1439x \]
b. \( \text{SST} = 52,120,800 \quad \text{SSE} = 7,102,922.54 \)
\[
\text{SSR} = \text{SST} - \text{SSR} = 52,120,800 - 7,102,922.54 = 45,017,877.46
\]
\[
\text{The estimated regression equation provided a very good fit.}
\]
c. \( \hat{y} = 28,574 - 1439x = 28,574 - 1439(15) = 6989 \)

Thus, an estimate of the price for a bike that weighs 15 pounds is $6989.

22 a. \( \text{SSE} = 1043.03 \)
\[
\bar{y} = \frac{\sum y_i}{n} = 677 \quad \text{SST} = \sum (y_i - \bar{y})^2 = 10,568
\]
\[
\text{SSR} = \text{SST} - \text{SSR} = 10,568 - 1043.03 = 9524.97
\]
\[
\text{The estimated regression equation provided a very good fit; approximately 90% of the variability in}
\]
the dependent variable was explained by the linear relationship between the two variables.

c. \( r = \sqrt{r^2} = \sqrt{.9013} = .95 \)

This reflects a strong linear relationship between the two variables.

23 a. \( s^2 = \text{MSE} = \frac{\text{SSE}}{n-2} = \frac{12.4}{3} = 4.133 \)

b. \( s = \sqrt{\text{MSE}} = \sqrt{4.133} = 2.033 \)

c. \( \Sigma (x_i - \bar{x})^2 = 10 \)
\[
\text{s_h} = \frac{s}{\sqrt{\Sigma (x_i - \bar{x})^2}} = \frac{2.033}{\sqrt{10}} = 0.643
\]
d. \( t = \frac{b_1}{s_h} = \frac{2.6}{.643} = 4.044 \)

Degrees of freedom = \( n - 2 = 3 \)
Because \( t > 0 \), \( p \)-value is two times the upper tail area

Using \( t \) table; area in upper tail is between .01 and .025; therefore, \( p \)-value is between .02 and .05.
Using Excel: \( p \)-value = 2*(1-DIST(4.044,3,TRUE)) = .0272
Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0272
Because \( p \)-value \( \leq \alpha \), we reject \( H_0: \beta_1 = 0 \)
e.  

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>67.6</td>
<td>1</td>
<td>67.6</td>
<td>16.36</td>
<td>.0272</td>
</tr>
<tr>
<td>Error</td>
<td>12.4</td>
<td>3</td>
<td>4.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>80.0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
MSR = SSR / 1 = 67.6
\]

\[
F = MSR / MSE = 67.6 / 4.133 = 16.36
\]

Using \( F \) table (1 degree of freedom numerator and 3 denominator), \( p \)-value is between .025 and .05

Using Excel, the \( p \)-value = F.DIST.RT(16.36,1,3) = .0272

Because \( p \)-value \( \leq \alpha \), we reject \( H_0: \beta_1 = 0 \)

24. a.  
\[
s^2 = MSE = SSE/(n - 2) = 230/3 = 76.6667
\]

b.  
\[
s = \sqrt{MSE} = \sqrt{76.6667} = 8.7560
\]

c.  
\[
\sum(x_i - \bar{x})^2 = 180
\]

\[
s_h = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} = \frac{8.7560}{\sqrt{180}} = 0.6526
\]

d.  
\[
t = \frac{b_1}{s_h} = \frac{-3}{0.6526} = -4.60
\]

Degrees of freedom = \( n - 2 = 3 \)

Because \( t < 0 \), \( p \)-value is two times the lower tail area

Using \( t \) table: area in lower tail is between .005 and .01; therefore, \( p \)-value is between .01 and .02.

Using Excel: \( p \)-value = 2*T.DIST(-4.60,3,TRUE) = .0193

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0193

Because \( p \)-value \( \leq \alpha \), we reject \( H_0: \beta_1 = 0 \)

e.  

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1620</td>
<td>1</td>
<td>1620</td>
<td>21.13</td>
<td>.0193</td>
</tr>
<tr>
<td>Error</td>
<td>230</td>
<td>3</td>
<td>76.6667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1850</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
MSR = SSR/1 = 1620
\]

\[
F = MSR/MSE = 1620/76.6667 = 21.13
\]

Using \( F \) table (1 degree of freedom numerator and 3 denominator), \( p \)-value is between .01 and .025.

Using Excel, the \( p \)-value = F.DIST.RT(21.13,1,3) = .0193.

Because \( p \)-value \( \leq \alpha \), we reject \( H_0: \beta_1 = 0 \)
26. a. In the statement of exercise 18, \( \hat{y} = 23.194 + .318x \)

In solving exercise 18, we found SSE = 287.624

\[
s^2 = \text{MSE} = \frac{\text{SSE}}{(n-2)} = \frac{287.624}{4} = 71.906
\]

\[
s = \sqrt{\text{MSE}} = \sqrt{71.906} = 8.4797
\]

\[
\sum (x - \bar{x})^2 = 14,950
\]

\[
s_h = \frac{s}{\sqrt{\sum (x - \bar{x})^2}} = \frac{8.4797}{\sqrt{14,950}} = 0.06935
\]

\[
t = \frac{b_1}{s_h} = \frac{.318}{0.06935} = 4.59
\]

Degrees of freedom = \( n - 2 = 4 \)

Because \( t > 0 \), \( p \)-value is two times the upper tail area

Using \( t \) table; area in upper tail is between .005 and .01; therefore, \( p \)-value is between .01 and .02.

Using Excel: \( p \)-value = 2*(1-T.DIST(4.59,4,TRUE)) = .0101

Using unrounded Test Statistic via Excel with cell referencing, \( p \)-value = .0101

Because \( p \)-value \( \leq \alpha \), we reject \( H_0: \beta_1 = 0 \); there is a significant relationship between price and overall score

b. In exercise 18 we found SSR = 1512.376

\[
\text{MSR} = \frac{\text{SSR}}{1} = 1512.376/1 = 1512.376
\]

\[
F = \frac{\text{MSR}}{\text{MSE}} = 1512.376/71.906 = 21.03
\]

Using \( F \) table (1 degree of freedom numerator and 4 denominator), \( p \)-value is between .01 and .025

Using Excel, the \( p \)-value = F.DIST.RT(21.03,1,4) = .0101.

Because \( p \)-value \( \leq \alpha \), we reject \( H_0: \beta_1 = 0 \)

c.  

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( F )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1512.376</td>
<td>1</td>
<td>1512.376</td>
<td>21.03</td>
<td>.0101</td>
</tr>
<tr>
<td>Error</td>
<td>287.624</td>
<td>4</td>
<td>71.906</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1800</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

28. The sum of squares due to error and the total sum of squares are

\[
\text{SSE} = \sum(y_i - \hat{y}_i)^2 = 1.4378 \quad \text{SST} = \sum(y_i - \bar{y})^2 = 3.5800
\]

Thus, \( \text{SSR} = \text{SST} - \text{SSE} = 3.5800 - 1.4378 = 2.1422 \)

\[
s^2 = \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{1.4378}{9} = .1598
\]

\[
s = \sqrt{\text{MSE}} = \sqrt{.1598} = .3997
\]
We can use either the $t$ test or $F$ test to determine whether speed of execution and overall satisfaction are related.

We will first illustrate the use of the $t$ test.

$$\Sigma(x_i - \bar{x})^2 = 2.6$$

$$s_h = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{3997}{\sqrt{2.6}} = 2.479$$

$$t = \frac{b_1}{s} = \frac{0.9077}{2.479} = 0.366$$

Degrees of freedom = $n - 2 = 9$

Because $t > 0$, $p$-value is two times the upper tail area

Using $t$ table; area in upper tail is less than .005; therefore, $p$-value is less than .01.

Using Excel: $p$-value = $2\times$1-T.DIST(3.66,9,TRUE)) = .0052

Using unrounded Test Statistic via Excel with cell referencing, $p$-value = .0052

Because $p$-value ≤ $\alpha$, we reject $H_0$: $\beta_1 = 0$

Because we can reject $H_0$: $\beta_1 = 0$ we conclude that speed of execution and overall satisfaction are related.

Next we illustrate the use of the $F$ test.

$$MSR = \frac{SSR}{1} = 2.1421$$

$$F = \frac{MSR}{MSE} = \frac{2.1421}{.1598} = 13.4$$

Using $F$ table (1 degree of freedom numerator and 9 denominator), $p$-value is less than .01

Using Excel, the $p$-value = F.DIST.RT(13.4,1,9) = .0052

Because $p$-value ≤ $\alpha$, we reject $H_0$: $\beta_1 = 0$

Because we can reject $H_0$: $\beta_1 = 0$ we conclude that speed of execution and overall satisfaction are related.

The ANOVA table is shown below.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2.1422</td>
<td>1</td>
<td>2.1422</td>
<td>13.4</td>
<td>.0052</td>
</tr>
<tr>
<td>Error</td>
<td>1.4378</td>
<td>9</td>
<td>.1598</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.5800</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

30. $SSE = \Sigma(y_i - \hat{y}_i)^2 = 1043.03$  $SST = \Sigma(y_i - \bar{y})^2 = 10,568$

Thus, $SSR = SST - SSE = 10,568 - 1043.03 = 9524.97$

$s^2 = MSE = SSE/(n-2) = 1043.03/4 = 260.758$
Chapter 14

$$s = \sqrt{260.758} = 16.1480$$

$$\sum (x_i - \bar{x})^2 = 56.655$$

$$s_h = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{16.148}{\sqrt{56.655}} = 2.145$$

$$t = \frac{\hat{b}}{s_h} = \frac{12.966}{2.145} = 6.04$$

Degrees of freedom = \(n - 2 = 44\)
Because \(t > 0\), \(p\)-value is two times the upper tail area

Using \(t\) table; area in upper tail is less than .005; therefore, \(p\)-value is less than .01.
Using Excel: \(p\)-value = 2*(1-T.DIST(6.04,4,TRUE)) = .0038
Using unrounded Test Statistic via Excel with cell referencing, \(p\)-value = .0038

Because \(p\)-value ≤ \(\alpha\), we reject \(H_0: \beta_1 = 0\)

There is a significant relationship between cars in service and annual revenue.

We can use either the \(t\) test or \(F\) test to determine whether variables are related. The solution for the \(F\) test is as follows:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>(F)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>9524.97</td>
<td>1</td>
<td>9524.97</td>
<td>36.53</td>
<td>.0038</td>
</tr>
<tr>
<td>Error</td>
<td>1043.03</td>
<td>4</td>
<td>260.758</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10568</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(SSE = \Sigma (y_i - \hat{y}_i)^2 = 1043.03 \quad SST = \Sigma (y_i - \bar{y})^2 = 10,568\)

Thus, \(SSR = SST - SSE = 10,568 - 1043.03 = 9524.97\)

\(s^2 = MSE = SSE/(n-2) = 1043.03/4 = 260.758\)

\(MSR = SSR/1 = 9524.97\)

\(F = MSR / MSE = 9524.97 / 260.758 = 36.53\)

Using \(F\) table (1 degree of freedom numerator and 4 denominator), \(p\)-value is less than .01
Using Excel, the \(p\)-value = F.DIST.RT(36.53,1,4) = .0038

Because \(p\)-value ≤ \(\alpha\), we reject \(H_0: \beta_1 = 0\). Cars in service and annual revenue are related.

32. a. \(SSE = \Sigma (y_i - \hat{y}_i)^2 = 12.4\)

\(s^2 = MSE = SSE / (n - 2) = 12.4 / 3 = 4.133\)

\(s = \sqrt{MSE} = \sqrt{4.133} = 2.033\)
\( \bar{x} = \frac{\sum x_i}{n} = \frac{15}{5} = 3 \)

\( \Sigma (x_i - \bar{x})^2 = 10 \)

\[ s_y = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 2.033\sqrt{\frac{1}{5} + \frac{(4 - 3)^2}{10}} = 1.114 \]

b. \( \hat{y}^* = .2 + 2.6 x^* = .2 + 2.6(4) = 10.6 \)

\( \hat{y}^* \pm t_{u/2} s_y \)  
\( \text{df} = n - 2 = 3 \quad t_{u/2} = 3.182 \)

\( 10.6 \pm 3.182 (1.114) = 10.6 \pm 3.54 \)

or 7.06 to 14.14

c. \( s_{pred} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 2.033\sqrt{1 + \frac{1}{5} + \frac{(4 - 3)^2}{10}} = 2.32 \)

d. \( \hat{y}^* \pm t_{u/2} s_{pred} \)  
\( \text{df} = n - 2 = 3 \quad t_{u/2} = 3.182 \)

\( 10.6 \pm 3.182 (2.32) = 10.6 \pm 7.38 \)

or 3.22 to 17.98

34. \( \text{SSE} = \sum (y_i - \hat{y}_i)^2 = 127.3 \)

\( s^2 = \text{MSE} = \text{SSE} / (n - 2) = 127.3 / 3 = 42.433 \)

\( s = \sqrt{\text{MSE}} = \sqrt{42.433} = 6.514 \)

\( \bar{x} = \frac{\sum x_i}{n} = \frac{50}{5} = 10 \)

\( \Sigma (x_i - \bar{x})^2 = 190 \)

\[ s_y = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 6.514\sqrt{\frac{1}{5} + \frac{(12 - 10)^2}{190}} = 3.063 \]

\( \hat{y}^* = 7.6 + .9 x^* = 7.6 + .9(12) = 18.40 \)

\( \hat{y}^* \pm t_{u/2} s_y \)  
\( \text{df} = n - 2 = 3 \quad t_{u/2} = 3.182 \)

\( 18.40 \pm 3.182(3.063) = 18.40 \pm 9.75 \)

or 8.65 to 28.15

\[ s_{pred} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 6.514\sqrt{1 + \frac{1}{5} + \frac{(12 - 10)^2}{190}} = 7.198 \]
\[
\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}} \quad \text{df} = n - 2 = 3 \quad t_{a/2} = 3.182
\]

\[
18.40 \pm 3.182(7.198) = 18.40 \pm 22.90
\]

or -4.50 to 41.30

The two intervals are different because there is more variability associated with predicting an individual value than there is a mean value.

35. a. \[
\hat{y}^* = 2090.5 + 581.1x^* = 2090.5 + 581.1(3) = 3833.8
\]

b. 
\[
s = \sqrt{\text{MSE}} = \sqrt{21,284} = 145.89
\]

\[
\bar{x} = \frac{\Sigma x_i}{n} = 19.2 / 6 = 3.2 \quad \Sigma (x_i - \bar{x})^2 = 0.74
\]

\[
s_{y^*} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 145.89 \sqrt{\frac{1}{6} + \frac{(3 - 3.2)^2}{0.74}} = 68.54
\]

\[
\hat{y}^* \pm t_{\alpha/2} s_{y^*} \quad \text{df} = n - 2 = 4 \quad t_{a/2} = 2.776
\]

\[
3833.8 \pm 2.776(68.54) = 3833.8 \pm 190.27
\]

or $3643.53$ to $4024.07$

or $3643.52$ to $4024.05$ from unrounded values and actual equation calculated from the data.

c. \[
s_{\text{pred}} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 145.89 \sqrt{1 + \frac{1}{6} + \frac{(3 - 3.2)^2}{0.74}} = 161.19
\]

\[
\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}} \quad \text{df} = n - 2 = 4 \quad t_{a/2} = 2.776
\]

\[
3833.8 \pm 2.776(161.19) = 3833.8 \pm 447.46
\]

or $3386.34$ to $4281.26$

or $3386.33$ to $4281.24$ from unrounded values and actual equation calculated from the data.

d. As expected, the prediction interval is much wider than the confidence interval. This is due to the fact that it is more difficult to predict the starting salary for one new student with a GPA of 3.0 than it is to estimate the mean for all students with a GPA of 3.0.

36. a. \[
s_{y^*} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma (x_i - \bar{x})^2}} = 4.6098 \sqrt{\frac{1}{10} + \frac{(9 - 7)^2}{142}} = 1.6503
\]

\[
\hat{y}^* \pm t_{\alpha/2} s_{y^*} \quad \text{df} = n - 2 = 8 \quad t_{a/2} = 2.306
\]

\[
\hat{y}^* = 80 + 4x^* = 80 + 4(9) = 116
\]

\[
116 \pm 2.306(1.6503) = 116 \pm 3.806
\]

or $112.194$ to $119.806$ ($112,194$ to $119,806$)
b. $s_{\text{pred}} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 4.6098 \sqrt{\frac{1}{10} + \frac{(9 - 7)^2}{142}} = 4.8963$

$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}} \text{ df } = n - 2 = 8 \quad t_{0.025} = 2.306$

$116 \pm 2.306(4.8963) = 116 \pm 11.291$

or $104.709$ to $127.291$ ($104,709$ to $127,291$)

c. As expected, the prediction interval is much wider than the confidence interval. This is due to the fact that it is more difficult to predict annual sales for one new salesperson with 9 years of experience than it is to estimate the mean annual sales for all salespersons with 9 years of experience.

38. a. $\hat{y} = 1246.67 + 7.6(500) = 5046.67$

b. $\text{SSE} = \sum(y_i - \hat{y})^2 = 233,333.33$

$s^2 = \text{MSE} = \text{SSE} / (n - 2) = 233333.33 / 4 = 58333.33$

$s = \sqrt{\text{MSE}} = \sqrt{58333.33} = 241.52$

$\bar{x} = \sum x_i / n = 3450 / 6 = 575$

$\sum(x_i - \bar{x})^2 = 93750$

$s_{\text{pred}} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 241.52 \sqrt{\frac{1}{6} + \frac{(500 - 575)^2}{93,750}} = 267.50$

$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}} \text{ df } = n - 2 = 4 \quad t_{0.025} = 4.604$

$5046.67 \pm 4.604(267.50) = 5046.67 \pm 1231.57 \quad \text{(or } 5046.667 \pm 1231.564 \text{ from unrounded)}$

or $3815.10$ to $6278.24 \quad \text{(or } 3815.10 \text{ to } 6278.23 \text{ from unrounded)}$

c. Based on one month, $6000$ is not out of line since $3815.10$ to $6278.24$ is the prediction interval. However, a sequence of five to seven months with consistently high costs should cause concern.

40. a. Total df = n-1. Therefore 8=n-1 and n=9

b. $\hat{y} = 20.0 + 7.21x$

c. Using the t stat for $\beta_1$ (coefficient of x variable), $t = 5.29$

Degrees of freedom = n – 2 = 7
Because t>0, p-value is two times the upper tail area

Using t table; area in upper tail is less than .005; therefore, p-value is less than .01.
Using Excel: p-value = 2*(1-T.DIST(5.29,7,TRUE)) = .0011

Because p-value ≤ α, we reject $H_0$: $\beta_1 = 0$
d. $SSE = SST - SSR = 51,984.1 - 41,587.3 = 10,396.8$

$MSE = SSE/(n-2) = 10,396.8 / 7 = 1,485.26$

$MSR = SSR/1 = 41587.26$

$F = MSR / MSE = 41,587.3 / 1,485.3 = 28.00$

From the $F$ table (1 degree of freedom numerator and 7 denominator), $p$-value is less than .01

Using Excel, $p$-value = F.DIST.RT(28,1,7) = .0011

Because $p$-value $\leq \alpha$, we reject $H_0: \beta_1 = 0$.

e. $\hat{y} = 20.0 + 7.21x = 20.0 + 7.21(50) = 380.5$ or $\$380,500$

42. a. $\hat{y} = 80.0 + 50.0x$

b. $SSE = SST - SSR = 9127.4 - 6826.6 = 2300.8$

$MSE = SSE/(n-2) = 2300.8 / 28 = 82.17$

$MSR = SSR/1 = 6828.6$

$F = MSR / MSE = 6828.6 / 82.17 = 83.1$

From the $F$ table (1 degree of freedom numerator and 28 denominator), $p$-value is less than .01

Using Excel, $p$-value = F.DIST.RT(83.1,1,28) = 0

Because $p$-value $\leq \alpha$, we reject $H_0: \beta_1 = 0$.

Branch office sales are related to the salespersons.

c. $t = \frac{50}{5.482} = 9.12$

Degrees of freedom = $n - 2 = 28$

Because $t > 0$, $p$-value is two times the upper tail area

Using $t$ table; area in upper tail is less than .005; therefore, $p$-value is less than .01.

Using Excel: $p$-value = 2*(1-T.DIST(9.12,28,TRUE)) = 0

Because $p$-value $\leq \alpha$, we reject $H_0: \beta_1 = 0$

d. $\hat{y}^* = 80 + 50x^* = 80 + 50(12) = 680$ or $\$680,000$
44. a. Scatter diagram:

![Scatter diagram showing a negative linear relationship between price and weight.](scatter_diagram.png)

b. There appears to be a negative linear relationship between the two variables. The heavier helmets tend to be less expensive.

c. Using datafile, RaceHelmets and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2044.3809</td>
<td>226.3543</td>
<td>9.0318</td>
<td>1.111E-07</td>
<td>1564.5313</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 2044.38 - 28.35 \text{ Weight} \]

d. From the Excel output for both the F test and the t test on \( \beta_1 \) (coefficient of x), there is evidence of a significant relationship: \( p-value = .000 < \alpha = .05 \)

e. \( r^2 = 0.774; \text{ A good fit} \)
46. a. Using Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

**SUMMARY OUTPUT**

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.620164219</td>
</tr>
<tr>
<td>R Square</td>
<td>0.384603659</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.296689895</td>
</tr>
<tr>
<td>Standard Error</td>
<td>2.054229935</td>
</tr>
<tr>
<td>Observations</td>
<td>9</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>18.46097561</td>
<td>18.46098</td>
<td>4.374783</td>
<td>0.074793318</td>
</tr>
<tr>
<td>Residual</td>
<td>7</td>
<td>29.53902439</td>
<td>4.219861</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                      |        |        |      |         |                |
| Coefficients        | Standard Error | t Stat | P-value |
| Intercept           | 2.32195122  | 1.88710113 | 1.230433 | 0.258275  |
| X Variable 1        | 0.636585366 | 0.304353556 | 2.091598 | 0.074793  |

\[ \hat{y} = 2.32 + 0.64x \]

b. From Excel’s Data Analysis Regression Tool using the residual output:

**RESIDUAL OUTPUT**

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted Y</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.595121951</td>
<td>0.404878049</td>
</tr>
<tr>
<td>2</td>
<td>4.231707317</td>
<td>0.768292683</td>
</tr>
<tr>
<td>3</td>
<td>4.868292683</td>
<td>-0.868292683</td>
</tr>
<tr>
<td>4</td>
<td>5.504878049</td>
<td>0.495121951</td>
</tr>
<tr>
<td>5</td>
<td>6.77804878</td>
<td>-2.77804878</td>
</tr>
<tr>
<td>6</td>
<td>6.77804878</td>
<td>-0.77804878</td>
</tr>
<tr>
<td>7</td>
<td>6.77804878</td>
<td>2.22195122</td>
</tr>
<tr>
<td>8</td>
<td>7.414634146</td>
<td>-2.414634146</td>
</tr>
<tr>
<td>9</td>
<td>8.051219512</td>
<td>2.948780488</td>
</tr>
</tbody>
</table>

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
The assumption that the variance is the same for all values of $x$ is questionable. The variance appears to increase for larger values of $x$.

48. Using Excel's Descriptive Statistics Regression Tool, the Excel output is shown below:

**SUMMARY OUTPUT**

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>t Stat</td>
</tr>
<tr>
<td>P-value</td>
</tr>
<tr>
<td>X Variable 1</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>t Stat</td>
</tr>
<tr>
<td>P-value</td>
</tr>
</tbody>
</table>

a. $\hat{y} = 80 + 4x$

**RESIDUAL OUTPUT**

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted Y</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84</td>
<td>-4</td>
</tr>
<tr>
<td>2</td>
<td>92</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>96</td>
<td>-4</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>104</td>
<td>-1</td>
</tr>
</tbody>
</table>
b. The assumptions concerning the error term appear reasonable.

50. a. The scatter diagram is shown below:

![Scatter Diagram]

The scatter diagram indicates that the first observation \((x = 135, y = 145)\) may be an outlier. For simple linear regression the scatter diagram can be used to check for possible outliers.
b. Using Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

### SUMMARY OUTPUT

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.620115502</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Square</td>
<td>0.384543235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.261451882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>12.61505192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>SS</td>
<td>MS</td>
<td>F</td>
<td>Significance F</td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>--------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>Regression</td>
<td>1</td>
<td>497.1594684</td>
<td>497.1594684</td>
<td>3.124047515</td>
</tr>
<tr>
<td>Residual</td>
<td>5</td>
<td>795.6976744</td>
<td>159.1395349</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>1292.857143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>66.10465116</td>
<td>32.06135318</td>
<td>2.061817254</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>0.402325581</td>
<td>0.22762441</td>
<td>1.767497529</td>
</tr>
</tbody>
</table>

Using equation 14.30 and 14.32 from the text to compute the standardized residual:

Residual = y - \( \hat{y} \)

\[ s^2 = MSE = 159.1395 \]

\[ s = \sqrt{MSE} = \sqrt{159.1395} = 12.6151 \]

\[ h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\Sigma(x_i - \bar{x})^2} = \frac{1}{7} + \frac{(x_i - 139.2857)^2}{3071.4286} = \text{leverage} \]

Standardized residual\(_i\) = \( \frac{\text{residual}_i}{s\sqrt{1-h_i}} \)

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>( \hat{y} )</th>
<th>y - ( \hat{y} )</th>
<th>residual</th>
<th>h</th>
<th>standardized residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>145</td>
<td>120.4186</td>
<td>24.5814</td>
<td>0.1488</td>
<td>0.1121</td>
<td>2.1121</td>
</tr>
<tr>
<td>110</td>
<td>100</td>
<td>110.3605</td>
<td>-10.3605</td>
<td>0.4221</td>
<td>1.0803</td>
<td>-1.0803</td>
</tr>
<tr>
<td>130</td>
<td>120</td>
<td>118.4070</td>
<td>1.5930</td>
<td>0.1709</td>
<td>0.1387</td>
<td>0.1387</td>
</tr>
<tr>
<td>145</td>
<td>120</td>
<td>124.4419</td>
<td>-4.4419</td>
<td>0.1535</td>
<td>-0.3827</td>
<td>-0.3827</td>
</tr>
<tr>
<td>175</td>
<td>130</td>
<td>136.5116</td>
<td>-6.5116</td>
<td>0.5581</td>
<td>-0.7765</td>
<td>-0.7765</td>
</tr>
<tr>
<td>160</td>
<td>110</td>
<td>114.3837</td>
<td>-4.3837</td>
<td>0.2826</td>
<td>-0.4046</td>
<td>-0.4046</td>
</tr>
<tr>
<td>120</td>
<td>110</td>
<td>114.3837</td>
<td>-4.3837</td>
<td>0.2640</td>
<td>-0.4050</td>
<td>-0.4050</td>
</tr>
</tbody>
</table>

Alternatively, as discussed in Section 14.8, Excel’s Descriptive Statistics Regression tool calculates a standard (vs standardized) residual that appears different (especially for smaller samples) but which will generally have little effect on the pattern observed.
The results of the standard residuals as calculated by Excel are as follows:

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted Y</th>
<th>Residuals</th>
<th>Standard Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120.4186047</td>
<td>24.58139535</td>
<td>2.13455875</td>
</tr>
<tr>
<td>2</td>
<td>110.3604651</td>
<td>-10.36046512</td>
<td>-0.899665017</td>
</tr>
<tr>
<td>3</td>
<td>118.4069767</td>
<td>1.593023256</td>
<td>0.138332331</td>
</tr>
<tr>
<td>4</td>
<td>124.4418605</td>
<td>-4.441860465</td>
<td>-0.385714968</td>
</tr>
<tr>
<td>5</td>
<td>136.5116279</td>
<td>-6.511627907</td>
<td>-0.565446027</td>
</tr>
<tr>
<td>6</td>
<td>130.4767442</td>
<td>-0.476744186</td>
<td>-0.041398727</td>
</tr>
<tr>
<td>7</td>
<td>114.3837209</td>
<td>-4.38372093</td>
<td>-0.380666343</td>
</tr>
</tbody>
</table>

Because the standard residual for the first observation is greater than 2 it is considered to be an outlier.

52. a.

The scatter diagram does indicate potential influential observations. For example, the 22.2% fundraising expense for the American Cancer Society and the 16.9% fundraising expense for the St. Jude Children’s Research Hospital look like they may each have a large influence on the slope of the estimated regression line. And, with a fundraising expense of on 2.6%, the percentage spent on programs and services by the Smithsonian Institution (73.7%) seems to be somewhat lower than would be expected; thus, this observation may need to be considered as a possible outlier.

b. Using datafile, Charities and Excel’s Descriptive Statistics Regression Tool, a portion of the Excel output follows:

<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>
\[ \hat{y} = 90.9815 - 0.9172 \text{ Fundraising Expenses (\%)} \]

c. The slope of the estimated regression equation is -0.9172. Thus, for every 1% increase in the amount spent on fundraising, the percentage spent on program expenses will decrease by .9172%; in other words, just a little under 1%. The negative slope and value seem to make sense in the context of this problem situation.

d. Using equation 14.30 and 14.32 from the text to compute the standardized residual:

Residual = \( y - \hat{y} \) 

\[ s^2 = MSE = 55.8587 \] 

\[ s = \sqrt{MSE} = \sqrt{55.8587} = 7.4739 \] 

\[ h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2} = \frac{1}{10} + \frac{(x_i - 6.26)^2}{485.444} = \text{ leverage} \] 

Standardized residual, \( r_i = \frac{\text{residual}}{s\sqrt{1 - h_i}} \)

Alternatively, as discussed in Section 14.8, Excel’s Descriptive Statistics Regression tool calculates a standard (vs standardized) residual that appears different (especially for smaller samples) but which will generally have little effect on the pattern observed.
The results of the standard residuals as calculated by Excel are as follows:

### RESIDUAL OUTPUT

<table>
<thead>
<tr>
<th>Observation</th>
<th>Predicted Y</th>
<th>Residuals</th>
<th>Standard Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87.49623479</td>
<td>4.603765213</td>
<td>0.653347356</td>
</tr>
<tr>
<td>2</td>
<td>84.10271092</td>
<td>4.19728908</td>
<td>0.595661941</td>
</tr>
<tr>
<td>3</td>
<td>88.59683712</td>
<td>-14.89683712</td>
<td>-2.114097633</td>
</tr>
<tr>
<td>4</td>
<td>88.78027084</td>
<td>8.019729155</td>
<td>1.138126858</td>
</tr>
<tr>
<td>5</td>
<td>70.62033231</td>
<td>0.979667686</td>
<td>0.139030394</td>
</tr>
<tr>
<td>6</td>
<td>89.23885515</td>
<td>0.16144849</td>
<td>0.022869012</td>
</tr>
<tr>
<td>7</td>
<td>89.51400573</td>
<td>-4.314005735</td>
<td>-0.61225887</td>
</tr>
<tr>
<td>8</td>
<td>90.33945749</td>
<td>8.460542514</td>
<td>1.20068527</td>
</tr>
<tr>
<td>9</td>
<td>75.48132596</td>
<td>-2.081325961</td>
<td>-0.295373189</td>
</tr>
<tr>
<td>10</td>
<td>88.22996968</td>
<td>-5.129969677</td>
<td>-0.728024121</td>
</tr>
</tbody>
</table>

The standardized residuals from the Excel output and the calculated values for leverage are shown below.

<table>
<thead>
<tr>
<th>Charity</th>
<th>Standard Residuals</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Red Cross</td>
<td>0.6533</td>
<td>0.1125</td>
</tr>
<tr>
<td>World Vision</td>
<td>0.5957</td>
<td>0.1032</td>
</tr>
<tr>
<td>Smithsonian Institution</td>
<td>-2.1141</td>
<td>0.1276</td>
</tr>
<tr>
<td>Food For The Poor</td>
<td>1.1381</td>
<td>0.1307</td>
</tr>
<tr>
<td>American Cancer Society</td>
<td>0.1390</td>
<td>0.6234</td>
</tr>
<tr>
<td>Volunteers of America</td>
<td>0.0229</td>
<td>0.1392</td>
</tr>
<tr>
<td>Dana-Farber Cancer Institute</td>
<td>-0.6122</td>
<td>0.1447</td>
</tr>
<tr>
<td>AmeriCares</td>
<td>1.2007</td>
<td>0.1637</td>
</tr>
<tr>
<td>ALSAC - St. Jude Children's Research Hospital</td>
<td>-0.2954</td>
<td>0.3332</td>
</tr>
<tr>
<td>City of Hope</td>
<td>-0.7280</td>
<td>0.1219</td>
</tr>
</tbody>
</table>

- Observation 3 (Smithsonian Institution) is considered to be an outlier because it has a large standardized residual; standard residual = -2.1141 < -2.

- Observation 5 (American Cancer Society) is an influential observation because it has high leverage; leverage = .6234 > 6/10.

Although fundraising expenses for the Smithsonian Institution are on the low side as compared to most of the other super-sized charities, the percentage spent on program expenses appears to be much lower than one would expect. It appears that the Smithsonian’s administrative expenses are too high. But, thinking about the expenses of running a large museum like the Smithsonian, the percentage spent on administrative expenses may not be unreasonable and is just due to the fact that operating costs for a museum are in general higher than for some other types of organizations. The very large value of fundraising expenses for the American Cancer Society suggests that this observation has a large influence on the estimated regression equation. The following Excel output shows the results if this observation is deleted from the original data.
Regression Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.5611</td>
</tr>
<tr>
<td>R Square</td>
<td>0.3149</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.2170</td>
</tr>
<tr>
<td>Standard Error</td>
<td>7.9671</td>
</tr>
<tr>
<td>Observations</td>
<td>9</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>204.1814</td>
<td>204.1814</td>
<td>3.2168</td>
<td>0.1160</td>
</tr>
<tr>
<td>Residual</td>
<td>7</td>
<td>444.3209</td>
<td>63.4744</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>648.5022</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.6537</td>
<td>24.9766</td>
<td>4.207E-08</td>
</tr>
<tr>
<td>Fundraising Expenses (%)</td>
<td>0.5590</td>
<td>-1.7935</td>
<td>0.1160</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 91.2561 - 1.0026 \text{ Fundraising Expenses (\%)} \]

The \( y \)-intercept has changed slightly, but the slope has changed from -0.917 to -1.0026.

a.

The scatter diagram does indicate potential outliers and/or influential observations. For example, the New York Yankees have both the highest revenue and value, and appears to be an influential observation. The Los Angeles Dodgers have the second highest value and appears to be an outlier.
b. Using datafile, MLBValues and Excel’s Descriptive Statistics Regression Tool, a portion of the Excel output follows:

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R                   0.9062</td>
</tr>
<tr>
<td>R Square                     0.8211</td>
</tr>
<tr>
<td>Adjusted R Square            0.8148</td>
</tr>
<tr>
<td>Standard Error               165.6581</td>
</tr>
<tr>
<td>Observations                 30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-601.4814</td>
<td>122.4288</td>
<td>-4.9129</td>
<td>3.519E-05</td>
<td>-852.2655</td>
</tr>
<tr>
<td>Revenue ($ millions)</td>
<td>5.9271</td>
<td>0.5228</td>
<td>11.3378</td>
<td>5.616E-12</td>
<td>4.8562</td>
</tr>
</tbody>
</table>

Thus, the estimated regression equation that can be used to predict the team’s value given the value of annual revenue is \( \hat{y} = -601.4814 + 5.9271 \text{ Revenue} \).

c. Using Excel’s Data Analysis Regression Tool with standardized residuals output, you find that the Standard Residual value for the Los Angeles Dodgers is 4.7 and should be treated as an outlier. To determine if the New York Yankees point is an influential observation we can remove the observation and compute a new estimated regression equation. The results show that the estimated regression equation is \( \hat{y} = -449.061 + 5.2122 \text{ Revenue} \). The following two scatter diagrams illustrate the small change in the estimated regression equation after removing the observation for the New York Yankees. These scatter diagrams show that the effect of the New York Yankees observation on the regression results is not that dramatic. (Note that leverage analysis will show that the leverage for the New York Yankees is \( h=0.63 \) which does exceed \( 6/n = 6/30 = .2 \) and indicate the possibility of it as an influential observation.

**Scatter Diagram Including the New York Yankees Observation**
Scatter Diagram Excluding the New York Yankees Observation

\[ y = 5.9271x - 601.48 \]
\[ R^2 = 0.8211 \]

\[ y = 5.2122x - 449.06 \]
\[ R^2 = 0.5888 \]
56. a. The scatter diagram suggests that there is a linear relationship between size and selling price and that as size increases, selling price increases.

b. Using datafile, WSHouses and Excel's Descriptive Statistics Regression Tool, the Excel output appears below:

The estimated regression equation is: $\hat{y} = -59.016 + 115.091x$

c. From the Excel output for both the F test and the t test on $\beta_1$ (coefficient of x), there is evidence of a significant relationship: $p$-value $= 0 \leq \alpha = .05$

d. $\hat{y} = -59.016 + 115.091$(square feet) = -59.016 + 115.091(2.0) = 171.167 or approximately $171,167.$

e. The estimated regression equation should provide a good estimate because $r^2 = 0.897.$

f. This estimated equation might not work well for other cities. Housing markets are also driven by other factors that influence demand for housing, such as job market and quality-of-life factors. For
example, because of the existence of high tech jobs and its proximity to the ocean, the house prices in Seattle, Washington might be very different from the house prices in Winston, Salem, North Carolina.

58. Using datafile, Jensen and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10.5280</td>
<td>3.7449</td>
<td>2.8113</td>
</tr>
<tr>
<td>Weekly Usage</td>
<td>0.9534</td>
<td>0.1382</td>
<td>6.9010</td>
</tr>
</tbody>
</table>

a. \( \hat{y} = 10.528 + .9534x \)

b. From the Excel output for both the F test and the t test on \( \beta_1 \) (coefficient of \( x \)), there is evidence of a significant relationship: \( p\text{-value} = .0001 < \alpha = .05 \), we reject \( H_0: \beta_1 = 0 \).

c. \( \text{SSE} = \Sigma (y_i - \hat{y}_i)^2 = 144.474 \)

\[ s^2 = \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{144.474}{8} = 18.059 \]

\[ s = \sqrt{\text{MSE}} = \sqrt{18.059} = 4.25 \]

\[ \overline{x} = \frac{\Sigma x_i}{n} = \frac{253}{10} = 25.3 \]

\[ \Sigma (x_i - \overline{x})^2 = 946.1 \]

\[ s_{pred} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{\Sigma (x_i - \overline{x})^2}} = 4.25 \sqrt{1 + \frac{1}{10} + \frac{(30 - 25.3)^2}{946.1}} = 4.504 \]

\[ \hat{y}^* = 10.528 + 0.9534(30) = 39.131 \]
39.131 ± 2.306(4.504) = 39.131 ± 10.386

The 95% prediction interval is 28.74 to 49.52 or $2874 to $4952

d. Yes, since the predicted expense for 30 hours is $3913. Therefore, a $3000 contract should save money.

60. a. Using datafile, HoursPts and Excel’s Descriptive Statistics Regression Tool, the Excel output follows:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.8470</td>
<td>7.9171</td>
<td>0.7335</td>
<td>-12.5358</td>
<td>24.2399</td>
<td>-12.5358</td>
<td>24.2399</td>
</tr>
<tr>
<td>Hours</td>
<td>0.8295</td>
<td>0.1095</td>
<td>7.5775</td>
<td>0.0001</td>
<td>0.5771</td>
<td>1.0820</td>
<td>0.5771</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 5.847 + .8295 \times \text{Hours} = 5.847 + .8295x \]

b. From the Excel output for both the F test and the t test on \( \beta_1 \) (coefficient of x), there is evidence of a significant relationship: \( p\text{-value} = .0001 < \alpha = .05 \), we reject \( H_0: \beta_1 = 0 \).

Total points earned is related to the hours spent studying.

c. Points = 5.8470 + .8295 Hours = 5.8470 + .8295(95) = 84.65 points

d. \( \text{SSE} = \sum (y_i - \hat{y}_i)^2 = 452.779 \)

\[ s^2 = \text{MSE} = \frac{\text{SSE}}{(n - 2)} = \frac{452.779}{8} = 56.597 \]

\[ s = \sqrt{\text{MSE}} = \sqrt{56.597} = 7.523 \]

\[ \bar{X} = \frac{\sum x_i}{n} = 695/10 = 69.5 \]

\[ \sum (x_i - \bar{X})^2 = 4722.5 \]

\[ \hat{y}_{\text{pred}} = \bar{y} = \frac{1}{n} + \frac{(x^* - \bar{X})^2}{\sum (x_i - \bar{X})^2} = 7.523 \left[ \frac{1}{10} + \frac{(95 - 69.5)^2}{4722.5} \right] = 8.370 \]

\[ \hat{y}^* \pm t_{\alpha/2} s_{\text{pred}} \text{ df} = n - 2 = 8 \quad t_{0.025} = 2.306 \]

\[ \hat{y}^* = 5.847 + .8295(95) = 84.653 \]

\[ 84.653 \pm 2.306(8.37) = 84.653 \pm 19.3 \]
The 95% prediction interval is 65.353 to 103.954

b. There appears to be a positive linear relationship between the two variables.

c. Using datafile, AirlineSeats and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below.

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>2011 Percentage</td>
</tr>
</tbody>
</table>

\[
\hat{y} = 7.3880 + 0.9276(2011 \text{ Percentage})
\]

d. From the Excel output for both the F test and the t test on \( \beta_1 \text{ (coefficient of x)} \), there is evidence of a significant relationship: \( p\)-value = 0.000 < \( \alpha = .05 \).

e. \( r^2 = .7572 \); a good fit.
The point with a residual value of approximately 36 clearly stands out as compared to the other points. This point corresponds to the observation for Air Tran Airways. Other than this point, the residual plot does not exhibit a pattern that would suggest a linear model is not appropriate.
Chapter 15
Multiple Regression

Solutions:

2. a. Using datafile Exer2 and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10021.24739</td>
<td>10021.25</td>
<td>15.5318</td>
<td>0.0043</td>
</tr>
<tr>
<td></td>
<td>5161.652607</td>
<td>645.2066</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>45.0594</td>
<td>25.4181</td>
<td>1.7727</td>
</tr>
<tr>
<td>X1</td>
<td>1.9436</td>
<td>0.4932</td>
<td>3.9410</td>
</tr>
</tbody>
</table>

An estimate of $y$ when $x_1 = 45$ is

$\hat{y} = 45.0594 + 1.9436(45) = 132.52$

b. Using datafile Exer2 and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>
Chapter 15

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>3363.4142</td>
<td>3363.414</td>
<td>2.2765</td>
<td>0.1698</td>
</tr>
<tr>
<td>Residual</td>
<td>8</td>
<td>11819.4858</td>
<td>1477.436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>15182.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<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>2</td>
<td>14052.15497</td>
<td>7026.077</td>
<td>43.4957</td>
<td>0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>7</td>
<td>1130.745026</td>
<td>161.535</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>15182.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<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>-18.3683</td>
<td>17.97150328</td>
<td>-1.0221</td>
<td>0.3408</td>
</tr>
<tr>
<td>X1</td>
<td>2.0102</td>
<td>0.2471</td>
<td>8.1345</td>
<td>8.19E-05</td>
</tr>
<tr>
<td>X2</td>
<td>4.7378</td>
<td>0.9484</td>
<td>4.9954</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

An estimate of y when x₂ = 15 is

\[ \hat{y} = 85.2171 + 4.3215(15) = 150.04 \]

c. Using datafile Exer2 and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

Regression Statistics

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.9620</td>
<td>R Square</td>
<td>0.9255</td>
</tr>
<tr>
<td>R Square</td>
<td>0.9255</td>
<td>Adjusted R Square</td>
<td>0.9042</td>
</tr>
<tr>
<td>Standard Error</td>
<td>12.7096</td>
<td>Observations</td>
<td>10</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>2</td>
<td>14052.15497</td>
<td>7026.077</td>
<td>43.4957</td>
<td>0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>7</td>
<td>1130.745026</td>
<td>161.535</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>15182.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<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>-18.3683</td>
<td>17.97150328</td>
<td>-1.0221</td>
<td>0.3408</td>
</tr>
<tr>
<td>X1</td>
<td>2.0102</td>
<td>0.2471</td>
<td>8.1345</td>
<td>8.19E-05</td>
</tr>
<tr>
<td>X2</td>
<td>4.7378</td>
<td>0.9484</td>
<td>4.9954</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

An estimate of y when x₁ = 45 and x₂ = 15 is

\[ \hat{y} = -18.3683 + 2.0102(45) + 4.7378(15) = 143.16 \]
4. a. \[ \hat{y} = 25 + 10(15) + 8(10) = 255; \text{ sales estimate: } $255,000 \]

b. Sales can be expected to increase by $10 for every dollar increase in inventory investment when advertising expenditure is held constant. Sales can be expected to increase by $8 for every dollar increase in advertising expenditure when inventory investment is held constant.

5. a. Using datafile Showtime and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

\begin{verbatim}
Regression Statistics
Multiple R          0.8078
R Square           0.6526
Adjusted R Square  0.5946
Standard Error     1.2152
Observations       8

ANOVA
\begin{tabular}{lcccr}
\hline
 & \textit{df} & \textit{SS} & \textit{MS} & \textit{F} & \textit{Significance F} \\
\hline
Regression       & 1   & 16.6401 & 16.6401 & 11.2688 & 0.0153 \\
Residual         & 6   & 8.8599  & 1.4767  &         &           \\
Total            & 7   & 25.5    &         &         &           \\
\hline
\end{tabular}

Coefficients
\begin{tabular}{lcccr}
\hline
 & \textit{Standard Error} & \textit{t Stat} & \textit{P-value} \\
\hline
Intercept        & 1.5824 & 56.0159 & 2.174E-09 \\
Television Advertising ($1000s) & 0.4778 & 3.3569 & 0.0153 \\
\hline
\end{tabular}

\[ \hat{y} = 88.6377 + 1.6039x_1 \]

where \( x_1 = \) television advertising ($1000s)

b. Using datafile Showtime and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

\begin{verbatim}
Regression Statistics
Multiple R          0.9587
R Square           0.9190
Adjusted R Square  0.8866
Standard Error     0.6426
Observations       8

ANOVA
\begin{tabular}{lcccr}
\hline
 & \textit{df} & \textit{SS} & \textit{MS} & \textit{F} & \textit{Significance F} \\
\hline
Regression       & 2   & 23.4354 & 11.7177 & 28.3778 & 0.0019 \\
Residual         & 5   & 2.0646  & 0.4129  &         &           \\
Total            & 7   & 25.5    &         &         &           \\
\hline
\end{tabular}

Coefficients
\begin{tabular}{lcccr}
\hline
 & \textit{Standard Error} & \textit{t Stat} & \textit{P-value} \\
\hline
Intercept        & 1.5739 & 52.8825 & 4.57E-08 \\
\hline
\end{tabular}
\end{verbatim}
Television Advertising ($1000s)   2.2902  0.3041  7.5319  0.0007
Newspaper Advertising ($1000s)  1.3010  0.3207  4.0567  0.0098

\[ \hat{y} = 83.2301 + 2.2902x_1 + 1.3010x_2 \]

where

- \( x_1 = \) television advertising ($1000s)
- \( x_2 = \) newspaper advertising ($1000s)

c. It is 1.6039 in part (a) and 2.2902 in part (b); in part (a) the coefficient is an estimate of the change in revenue due to a one-unit change in television advertising expenditures; in part (b) it represents an estimate of the change in revenue due to a one-unit change in television advertising expenditures with the amount of newspaper advertising is held constant.

d. Revenue = 83.2301 + 2.2902(3.5) + 1.3010(1.8) = $93.588 or $93,588

6. a. Using datafileNFLPassing and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-58.7703</td>
<td>26.1754</td>
<td>2.2452</td>
</tr>
<tr>
<td>Yds/Att</td>
<td>16.3906</td>
<td>3.7497</td>
<td>4.3712</td>
</tr>
</tbody>
</table>

\[ \hat{y} = -58.7703 + 16.3906\text{Yds/Att} \]

b. Using datafileNFLPassing and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>
</tbody>
</table>
Multiple Regression

Standard Error 18.3008
Observations 16

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>3652.8003</td>
<td>3652.8003</td>
<td>10.9065</td>
<td>0.0052</td>
</tr>
<tr>
<td>Residual</td>
<td>14</td>
<td>4688.8697</td>
<td>334.9193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>8341.67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>97.5383</td>
<td>7.0365</td>
<td>5.898E-06</td>
</tr>
<tr>
<td>Int/Att</td>
<td>-1600.491</td>
<td>-3.3025</td>
<td>0.0052</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 97.5383 - 1600.491\text{Int/Att} \]

c. Using datafileNFLPassing and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

Regression Statistics

Multiple R 0.8675
R Square 0.7525
Adjusted R Square 0.7144
Standard Error 12.6024
Observations 16

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>2</td>
<td>6277.0142</td>
<td>3138.5071</td>
<td>19.7614</td>
<td>0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>13</td>
<td>2064.6558</td>
<td>158.8197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>8341.67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5.7633</td>
<td>-0.2123</td>
<td>0.8352</td>
</tr>
<tr>
<td>Yds/Att</td>
<td>12.9494</td>
<td>4.0649</td>
<td>0.0013</td>
</tr>
<tr>
<td>Int/Att</td>
<td>-1083.7880</td>
<td>-3.0348</td>
<td>0.0096</td>
</tr>
</tbody>
</table>

\[ \hat{y} = -5.7633 + 12.9494\text{Yds/Att} - 1083.7880\text{Int/Att} \]

d. The predicted value of Win% for the Kansas City Chiefs is

Win% = -5.7633 + 12.9494(6.2) - 1083.7880(0.036) = 35.5%

With 7 wins and 9 loses, the Kansas City Chiefs won 43.75% of the games they played. The predicted value is somewhat lower than the actual value.
8. a. Using datafile Ships and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<p>| ANOVA |</p>
<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>1</td>
<td>60.2022</td>
<td>60.2022</td>
<td>17.2106</td>
</tr>
<tr>
<td>Residual</td>
<td>18</td>
<td>62.9633</td>
<td>3.4980</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>123.1655</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>69.2998</td>
<td>4.7995</td>
<td>14.4390</td>
</tr>
<tr>
<td>Shore Excursions</td>
<td>0.2348</td>
<td>0.0566</td>
<td>4.1486</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 69.2998 + 0.2348 \text{ Shore Excursions} \]

b. Using datafile Ships and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<p>| ANOVA |</p>
<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>2</td>
<td>90.9545</td>
<td>45.4773</td>
<td>24.0015</td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>32.2110</td>
<td>1.8948</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>123.1655</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>45.1780</td>
<td>6.9518</td>
<td>6.4987</td>
</tr>
<tr>
<td>Shore Excursions</td>
<td>0.2529</td>
<td>0.0419</td>
<td>6.0369</td>
</tr>
</tbody>
</table>
\[ \hat{y} = 45.1780 + 0.2529 \text{ Shore Excursions} + 0.2482 \text{ Food/Dining} \]

c. The predicted score is
\[ \hat{y} = 45.1780 + 0.2529(80) + 0.2482(90) = 87.75 \text{ or approximately 88} \]

10. a. Using datafileMLBPitching and Excel’s Descriptive Statistics Regression Tool, the Excel output follows.

```
Regression Statistics
Multiple R 0.6477
R Square 0.4195
Adjusted R Square 0.3873
Standard Error 0.0603
Observations 20

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>1</td>
<td>0.0473</td>
<td>0.0473</td>
<td>13.0099</td>
</tr>
<tr>
<td>Residual</td>
<td>18</td>
<td>0.0654</td>
<td>0.0036</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>0.1127</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients
<table>
<thead>
<tr>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0631</td>
<td>10.7135</td>
</tr>
<tr>
<td>SO/IP</td>
<td>0.0787</td>
<td>-3.6069</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 0.6758 - 0.2838 \text{ SO/IP} \]

b. Using datafileMLBPitching and Excel’s Descriptive Statistics Regression Tool, the Excel output follows.

```
Regression Statistics
Multiple R 0.5063
R Square 0.2563
Adjusted R Square 0.2150
Standard Error 0.0682
Observations 20

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>1</td>
<td>0.0289</td>
<td>0.0289</td>
<td>6.2035</td>
</tr>
</tbody>
</table>
```
\[ \hat{y} = 0.3081 + 1.3467 \text{ HR/IP} \]

e. Using datafileMLBPitching and Excel’s Descriptive Statistics Regression Tool, the Excel output follows.

\[
\begin{array}{l}
\text{Regression Statistics} \\
\text{Multiple R} & 0.7506 \\
\text{R Square} & 0.5635 \\
\text{Adjusted R Square} & 0.5121 \\
\text{Standard Error} & 0.0538 \\
\text{Observations} & 20 \\
\end{array}
\]

\[
\begin{array}{l}
\text{ANOVA} \\
\begin{array}{llllll}
\text{df} & \text{SS} & \text{MS} & \text{F} & \text{Significance F} \\
\hline
\text{Regression} & 2 & 0.0635 & 0.0317 & 10.9714 & 0.0009 \\
\text{Residual} & 17 & 0.0492 & 0.0029 & & \\
\text{Total} & 19 & 0.1127 & & &
\end{array}
\end{array}
\]

\[
\begin{array}{l}
\text{Coefficients} \\
\begin{array}{llllll}
\text{Intercept} & 0.5365 & 0.0814 & 6.5903 & 4.58698E-06 \\
\text{SO/IP} & -0.2483 & 0.0718 & -3.4586 & 0.0030 \\
\text{HR/IP} & 1.0319 & 0.4359 & 2.3674 & 0.0300 \\
\end{array}
\end{array}
\]

\[ \hat{y} = 0.5365 - 0.2483 \text{ SO/IP} + 1.0319 \text{ HR/IP} \]

d. Using the estimated regression equation in part (c) we obtain

\[ \text{R/IP} = 0.5365 - 0.2483 \text{ SO/IP} + 1.0319 \text{ HR/IP} \]
\[ \text{R/IP} = 0.5365 - 0.2483(0.91) + 1.0319(0.16) = 0.48 \]

The predicted value for R/IP was less than the actual value.

e. This suggestion does not make sense. If a pitcher gives up more runs per inning pitched this pitcher’s earned run average also has to increase. For these data the sample correlation coefficient between ERA and R/IP is .964.

12. a A portion of the Excel output for part (c) of exercise 2 is shown below.

\[
\begin{array}{l}
\text{Regression Statistics} \\
\text{Multiple R} & 0.9620 \\
\text{R Square} & 0.9255 \\
\end{array}
\]
Multiple Regression

Adjusted R Square: 0.9042
Standard Error: 12.7096
Observations: 10

\[ R^2 = \frac{SSR}{SST} = \frac{14,052.2}{15,182.9} = .9255 \]

b. \[ R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} = 1 - (1 - .9255) \frac{10-1}{10-2-1} = .9042 \]

c. Yes; after adjusting for the number of independent variables in the model, we see that 90.42% of the variability in \( y \) has been accounted for.

14. a. \[ R^2 = \frac{SSR}{SST} = \frac{12,000}{16,000} = .75 \]

b. \[ R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} = 1 - (1 - .75) \frac{10-1}{10-2-1} = .6786 \]

c. The adjusted coefficient of determination shows that 67.86% of the variability has been explained by the two independent variables; thus, we conclude that the model does not explain a large amount of variability.

15. a. A portion of the Excel output for part (b) of exercise 5 is shown below.

<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>

\[ R^2 = \frac{SSR}{SST} = \frac{23.435}{25.5} = .9190 \]

\[ R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} = 1 - (1 - .9190) \frac{8-1}{8-2-1} = .8866 \]

b. Multiple regression analysis is preferred since both \( R^2 \) and \( R_a^2 \) show an increased percentage of the variability of \( y \) explained when both independent variables are used.

16. a. A portion of the Excel output for part (a) of exercise 6 is shown below.

<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>

\[ R^2 = .5771 \]. Thus, the averages number of passing yards per attempt is able to explain 57.71% of the
variability in the percentage of games won. Considering the nature of the data and all the other factors that might be related to the number of games won, this is not too bad a fit.

b. A portion of the Excel output for part (c) of exercise 6 is shown below.

<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>

The value of the coefficient of determination increased to $R^2 = .7525$, and the adjusted coefficient of determination is $R_a^2 = .7144$. Thus, using both independent variables provides a much better fit.

18. a. A portion of the Excel output for part (c) of exercise 10 is shown below.

<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>

The value of $R^2 = .5635$ and the value of $R_a^2 = .5121$.

b. The fit is not great, but considering the nature of the data being able to explain slightly more than 50% of the variability in the number of runs given up per inning pitched using just two independent variables is not too bad.

c. Using datafileMLBPitching and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below.

<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>

<table>
<thead>
<tr>
<th>ANOVA</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>2</td>
<td>5.1739</td>
<td>2.5870</td>
<td>14.1750</td>
<td>0.0002</td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>3.1025</td>
<td>0.1825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>8.2765</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Coefficients</th>
<th>Intercept</th>
<th>SO/IP</th>
<th>HR/IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>3.8781</td>
<td>-1.8428</td>
<td>11.9933</td>
</tr>
<tr>
<td>Error</td>
<td>0.6466</td>
<td>0.5703</td>
<td>3.4621</td>
</tr>
<tr>
<td>t Stat</td>
<td>5.9976</td>
<td>-3.2310</td>
<td>3.4641</td>
</tr>
<tr>
<td>P-value</td>
<td>1.44078E-05</td>
<td>0.0049</td>
<td>0.0030</td>
</tr>
</tbody>
</table>
The Excel output shows that $R^2 = .6251$ and $R_a^2 = .5810$.

Approximately 60% of the variability in the ERA can be explained by the linear effect of HR/IP and SO/IP. This is not too bad considering the complexity of predicting pitching performance.

19. a. $\text{SSE} = \text{SST} - \text{SSR} = 6724.125 - 6216.375 = 507.75$

$\text{MSR} = \frac{\text{SSR}}{p} = \frac{6216.375}{2} = 3108.1875$

$\text{MSE} = \frac{\text{SSE}}{n - p - 1} = \frac{507.75}{10 - 2 - 1} = 72.5357$

b. $F = \frac{\text{MSR/MSE}}{} = \frac{3108.1875/72.5357}{} = 42.85$

Using $F$ Table (2 degrees of freedom numerator and 7 denominator), $p$-value is less than .01.

Using Excel, the $p$-value = F.DIST.RT(42.85,2,7) = .0001

Because the $p$-value $\leq \alpha = .05$, the overall model is significant.

c. $t = \frac{b_1}{s_{b_1}} = \frac{.5906}{.0813} = 7.26$

Degrees of freedom $= n - p - 1 = 7$

Because $t > 0$, $p$-value is two times the upper tail area

Using $t$ table; area in upper tail is less than .005; therefore, $p$-value is less than .01.

Using Excel: $p$-value = 2*(1-T.DIST(7.26,7,TRUE)) = .0002

Because the $p$-value $\leq \alpha = .05$, $\beta_1$ is significant.

d. $t = \frac{b_2}{s_{b_2}} = \frac{.4980}{.0567} = 8.78$

Degrees of freedom $= n - p - 1 = 7$

Because $t > 0$, $p$-value is two times the upper tail area

Using $t$ table; area in upper tail is less than .005; therefore, $p$-value is less than .01.

Using Excel: $p$-value = 2*(1-T.DIST(8.78,7,TRUE)) = .0001

Because the $p$-value $\leq \alpha = .05$, $\beta_2$ is significant.

20. A portion of the Excel output for part (c) of exercise 2 is shown below.

<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>2</td>
<td>14052.15497</td>
<td>7026.077</td>
<td>43.4957</td>
<td>0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>7</td>
<td>1130.745026</td>
<td>161.535</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>15182.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-18.36826758</td>
<td>-1.0221</td>
<td>0.3408</td>
</tr>
<tr>
<td>X1</td>
<td>2.0102</td>
<td>8.1345</td>
<td>8.19E-05</td>
</tr>
<tr>
<td>X2</td>
<td>4.7378</td>
<td>4.9954</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

a. Since the p-value corresponding to \( F = 43.4957 \) is \( .0001 < \alpha = .05 \), we reject \( H_0: \beta_1 = \beta_2 = 0 \); there is a significant relationship.

b. For X1, since the p-value corresponding to \( t = 8.1345 \) is \( .0001 < \alpha = .05 \), we reject \( H_0: \beta_1 = 0 \); \( \beta_1 \) is significant.

c. For X2, since the p-value corresponding to \( t = 4.9954 \) is \( .0016 < \alpha = .05 \), we reject \( H_0: \beta_2 = 0 \); \( \beta_2 \) is significant.

22. a. \[ SSE = SST - SSR = 16000 - 12000 = 4000 \]

\[ MSE = \frac{SSE}{n - p - 1} = \frac{4000}{7} = 571.4286 \]

\[ MSR = \frac{SSR}{p} = \frac{12000}{2} = 6000 \]

b. \[ F = MSR/MSE = 6000/571.4286 = 10.50 \]

Using \( F \) Table (2 degrees of freedom numerator and 7 denominator), p-value is less than .01.

Using Excel, the p-value = F.DIST.RT(10.50,2,7) = .0078
Because the p-value \( \leq \alpha = .05 \), we reject \( H_0 \). There is a significant relationship among the variables.

23. A portion of the Excel output for part (c) of exercise 2 is shown below.

| Multiple R | 0.9587 |
| R Square   | 0.9190 |
| Adjusted R Square | 0.8866 |
| Standard Error | 0.6426 |
| Observations | 8 |

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>2</td>
<td>23.4354</td>
<td>11.7177</td>
<td>28.3778</td>
<td>0.0019</td>
</tr>
<tr>
<td>Residual</td>
<td>5</td>
<td>2.0646</td>
<td>0.4129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>25.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Coefficients | Standard Error | t Stat | P-value
--- | --- | --- | ---
Intercept | 83.2301 | 1.5739 | 52.8825 | 4.57E-08
Television Advertising ($1000s) | 2.2902 | 0.3041 | 7.5319 | 0.0007
Newspaper Advertising ($1000s) | 1.3010 | 0.3207 | 4.0567 | 0.0098

a. Since the p-value corresponding to $F = 28.3778$ is $0.0019 < \alpha = 0.01$, we reject $H_0: \beta_1 = \beta_2 = 0$; there is a significant relationship.

b. For Television Advertising, since the p-value corresponding to $t = 7.5319$ is $0.0007 < \alpha = 0.05$, we reject $H_0: \beta_1 = 0$; $\beta_1$ is significant and $x_1$ should not be dropped from the model.

c. For Newspaper Advertising, since the p-value corresponding to $t = 4.0567$ is $0.0098 < \alpha = 0.05$, we reject $H_0: \beta_2 = 0$; $\beta_2$ is significant and $x_2$ should not be dropped from the model.

24. a. Using datafile NFL2011 and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

Regression Statistics

| Multiple R | 0.6901 |
| R Square | 0.4762 |
| Adjusted R Square | 0.4401 |
| Standard Error | 15.3096 |
| Observations | 32 |

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>2</td>
<td>6179.1015</td>
<td>3089.5507</td>
<td>13.1815</td>
</tr>
<tr>
<td>Residual</td>
<td>29</td>
<td>6797.1673</td>
<td>234.3851</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>12976.2688</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients | Standard Error | t Stat | P-value
--- | --- | --- | ---
Intercept | 60.5405 | 28.3562 | 2.1350 | 0.0413 |
OffPassYds/G | 0.3186 | 0.0626 | 5.0929 | 1.95917E-05 |
DefYds/G | -0.2413 | 0.0893 | -2.7031 | 0.0114 |

\[ \hat{y} = 60.5405 + 0.3186 \text{ OffPassYds/G} - 0.2413 \text{ DefYds/G} \]

b. With $F=13.1815$, the p-value for the $F$ test $= .0001 < \alpha = .05$, there is a significant relationship.

c. For OffPassYds/G, $t=5.0929$: Because the p-value $= .0000 < \alpha = .05$, OffPassYds/G is significant.

For DefYds/G, $t=-2.7031$: Because the p-value $= .0114 < \alpha = .05$, DefYds/G is significant.

26. The Excel output from part (c) of exercise 10 follows:

Regression Statistics

| Multiple R | 0.7506 |
| R Square | 0.5635 |
Adjusted R Square 0.5121
Standard Error 0.0538
Observations 20

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>2</td>
<td>0.0635</td>
<td>0.0317</td>
<td>10.9714</td>
<td>0.0009</td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>0.0492</td>
<td>0.0029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>0.1127</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.5365</td>
<td>6.5903</td>
<td>4.5869E-06</td>
</tr>
<tr>
<td>SO/IP</td>
<td>-0.2483</td>
<td>-3.4586</td>
<td>0.0030</td>
</tr>
<tr>
<td>HR/IP</td>
<td>1.0319</td>
<td>2.3674</td>
<td>0.0300</td>
</tr>
</tbody>
</table>

a. The p-value associated with $F = 10.9714$ is .0009. Because the p-value $\leq .05$, there is a significant overall relationship.

b. For SO/IP, the p-value associated with $t = -3.4586$ is .0030. Because the p-value $\leq .05$, SO/IP is significant. For HR/IP, the p-value associated with $t = 2.3674$ is .0300. Because the p-value $\leq .05$, HR/IP is also significant.

28. a. $\hat{y} = -18.4 + 2.01(45) + 4.74(15) = 143.15$ or 143.16 from unrounded equation values and StatTools output.

   b. Using StatTools with Exer2 datafile, the 95% prediction interval is 111.16 to 175.16.

29. a. $\hat{y} = 83.2 + 2.29(3.5) + 1.30(1.8) = 93.555$ or $93,555$ or $93.588$ or $93,588$ from unrounded equation values and StatTools output (also ties to exercise 5d)

   b. Using StatTools with Showtime datafile, the prediction interval estimate: $91.774$ to $95.401$ or $91,774$ to $95,401$

30. a. A portion of the Excel output from Exercise 24 is shown below:

<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>60.5405</td>
<td>28.3562</td>
<td>2.1350</td>
<td>0.0413</td>
</tr>
<tr>
<td>OffPassYds/G</td>
<td>0.3186</td>
<td>0.0626</td>
<td>5.0929</td>
<td>1.95917E-05</td>
</tr>
<tr>
<td>DefYds/G</td>
<td>-0.2413</td>
<td>0.0893</td>
<td>-2.7031</td>
<td>0.0114</td>
</tr>
</tbody>
</table>

The estimated regression equation is

$\hat{y} = 60.5405 + 0.3186 \text{OffPassYds/G} - 0.2413 \text{DefYds/G}$

For OffPassYds/G = 225 and DefYds/G = 300, the predicted value of the percentage of games won is $\hat{y} = 60.5405 + 0.3186(225) - 0.2413(300) = 59.827$
b. Using StatTools with NFL2011 datafile, the 95% prediction interval is 26.959 to 92.695 or 27.0 to 92.7

32. a. \( E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \) where \( x_2 = 0 \) if level 1 and 1 if level 2
b. \( E(y) = \beta_0 + \beta_1 x_1 + \beta_2(0) = \beta_0 + \beta_1 x_1 \)
c. \( E(y) = \beta_0 + \beta_1 x_1 + \beta_2(1) = \beta_0 + \beta_1 x_1 + \beta_2 \)
d. \( \beta_2 \) is the change in \( E(y) \) for a 1 unit change in \( x_1 \) holding \( x_2 \) constant.
\[ \beta_2 = E(y \mid \text{level 2}) - E(y \mid \text{level 1}) \]

34. a. $15,300 which equals the coefficient of variable 3
b. Estimate of sales = 10.1 - 4.2(2) + 6.8(8) + 15.3(0) = 56.1 or $56,100
c. Estimate of sales = 10.1 - 4.2(1) + 6.8(3) + 15.3(1) = 41.6 or $41,600

36. a. Using datafile Repair and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.8602</td>
<td>2.5529</td>
<td>0.0433</td>
</tr>
<tr>
<td>Months Since Last Service</td>
<td>0.2914</td>
<td>3.4862</td>
<td>0.0130</td>
</tr>
<tr>
<td>Type</td>
<td>1.1024</td>
<td>3.6342</td>
<td>0.0109</td>
</tr>
<tr>
<td>Person</td>
<td>-0.6091</td>
<td>-1.5701</td>
<td>0.1674</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 1.8602 + .2914 \text{ Months} + 1.1024 \text{ Type} - .6091 \text{ Person} \]

b. Since the \( p \)-value corresponding to \( F = 18.04 \) is \( .0021 < \alpha = .05 \), the overall model is statistically significant.
c. The \( p \)-value corresponding to \( t = -1.5701 \) is \( .1674 > \alpha = .05 \); thus, the addition of Person is not statistically significant. Person is highly correlated with Months (the sample correlation coefficient is -.691); thus, once the effect of Months has been accounted for, Person will not add much to the model.
38. a. Using datafile Stroke and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-91.7595</td>
<td>-6.0278</td>
<td>1.76E-05</td>
</tr>
<tr>
<td>Age</td>
<td>1.0767</td>
<td>6.4878</td>
<td>7.49E-06</td>
</tr>
<tr>
<td>Pressure</td>
<td>0.2518</td>
<td>5.5680</td>
<td>4.24E-05</td>
</tr>
<tr>
<td>Smoker</td>
<td>8.7399</td>
<td>2.9125</td>
<td>0.0102</td>
</tr>
</tbody>
</table>

\[ \hat{y} = -91.7595 + 1.0767 \text{ Age} + 0.2518 \text{ Pressure} + 8.7399 \text{ Smoker} \]

b. Since the \( p \)-value corresponding to \( t = 2.9125 \) is \( .0102 \leq \alpha = .05 \), smoking is a significant factor.

c. \[ \hat{y} = -91.7595 + 1.0767 (68) + 0.2518 (175) + 8.7399 (1) = 34.2661 \]

The point estimate is 34.27; the 95% prediction interval is 21.35 to 47.18. Thus, the probability of a stroke (.2135 to .4718 at the 95% confidence level) appears to be quite high. The physician would probably recommend that Art quit smoking and begin some type of treatment designed to reduce his blood pressure.

40. a. The Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

\[ \hat{y} = -91.7595 + 1.0767 \text{ Age} + 0.2518 \text{ Pressure} + 8.7399 \text{ Smoker} \]
Multiple Regression

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-53.28</td>
<td>5.7864</td>
<td>-9.2079</td>
</tr>
<tr>
<td>x</td>
<td>3.11</td>
<td>0.2016</td>
<td>15.4283</td>
</tr>
</tbody>
</table>

\[ \hat{y} = -53.28 + 3.11 \, x \]

b. Excel's standard residuals are shown below:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
<th>Standard Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>12</td>
<td>-1.27</td>
</tr>
<tr>
<td>24</td>
<td>21</td>
<td>-.15</td>
</tr>
<tr>
<td>26</td>
<td>31</td>
<td>1.39</td>
</tr>
<tr>
<td>28</td>
<td>35</td>
<td>.49</td>
</tr>
<tr>
<td>40</td>
<td>70</td>
<td>-.45</td>
</tr>
</tbody>
</table>

Because none of the standard residuals are less than -2 or greater than 2, none of the observations can be classified as an outlier.

c. The standardized residual plot follows:

With only five points it is difficult to determine if the model assumptions are violated. The conclusions reached in part (b) regarding outliers also apply here. But, the point corresponding to observation 5 does appear to be unusual. To investigate this further, consider the following scatter diagram.
The scatter diagram indicates that observation 5 is influential.

42. a. Using datafile Auto2 and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

<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>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>71.3283</td>
<td>2.2479</td>
<td>1.06E-13</td>
</tr>
<tr>
<td>Price ($1000s)</td>
<td>0.1072</td>
<td>0.0392</td>
<td>0.0170</td>
</tr>
<tr>
<td>Horsepower</td>
<td>0.0845</td>
<td>0.0093</td>
<td>5.45E-07</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 71.3283 + 0.1072 \text{ Price } + 0.0845 \text{ Horsepower} \]

b. The standardized residual plot is shown below. There appears to be a very unusual trend in the residual plot. A different model should be considered.
c. The standardized residual plot did not identify any observations with a large standardized residual; thus, there does not appear to be any outliers in the data.

44. a. The expected increase in final college grade point average corresponding to a one point increase in high school grade point average is .0235 when SAT mathematics score does not change. Similarly, the expected increase in final college grade point average corresponding to a one point increase in the SAT mathematics score is .00486 when the high school grade point average does not change.

b. \[ \hat{y} = -1.41 + 0.0235(84) + 0.00486(540) = 3.1884 \text{ or about 3.19} \]

46. a. The computer output with the missing values filled in is as follows:

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.9607</td>
</tr>
<tr>
<td>R Square</td>
<td>0.923</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.9102</td>
</tr>
<tr>
<td>Standard Error</td>
<td>3.35</td>
</tr>
<tr>
<td>Observations</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Significance F</td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>2</td>
</tr>
<tr>
<td>1612</td>
<td>806</td>
</tr>
<tr>
<td>71.92</td>
<td>.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>12</td>
</tr>
<tr>
<td>134.48</td>
<td>11.21</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
</tr>
<tr>
<td>1746.48</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.103</td>
<td>2.667</td>
<td>3.0382</td>
</tr>
<tr>
<td>X1</td>
<td>7.602</td>
<td>2.105</td>
<td>3.6114</td>
</tr>
<tr>
<td>X2</td>
<td>3.111</td>
<td>0.613</td>
<td>5.0750</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 8.103 + 7.602 X1 + 3.111 X2 \]
To start, we can calculate the Multiple R = \( r = \sqrt{R^2} \) with the same sign as the slope. Therefore, \( r = \sqrt{.923} = .9607 \)

Regression df = \( p \) = number of independent variables = 2
Regression df + Residual df = Total df = 2 + 12 = 14
Total df = \( n-1 \) = 14
Therefore \( n=15 \) (number of observations)

\[
R^2 = 1 - (1-R^2) = 1 - (1-.923) = \frac{15-1}{15-2-1} = .9102
\]

\[
R^2 = \frac{\text{SSR}}{\text{SST}} \\
\text{and therefore, } \frac{\text{SST}}{R^2} = \frac{1612}{.923} = 1746.479
\]

\[
\text{SSE} = \text{SST} - \text{SSR} = 1746.479 - 1612 = 34.479
\]

\[
\text{MSR} = \frac{\text{SSR}}{p} = \frac{1612}{2} = 806
\]

\[
\text{MSE} = \frac{\text{SSE}}{n-p-1} = \frac{1746.479}{15-2-1} = 11.207
\]

\[
F = \frac{\text{MSR}}{\text{MSE}} = \frac{806}{11.207} = 71.92
\]
Using Excel, the \( p \)-value = \( \text{F.DIST.RT}(71.92,2,12) = .0000 \)

\[
t = \frac{b}{s_{b_i}} \text{ and therefore, for the Intercept, } t = \frac{8.103}{2.667} = 3.0382
\]

For the X1 coefficient, \( t = \frac{7.602}{2.105} = 3.6114 \)

For the X2 coefficient, \( t = \frac{3.111}{.613} = 5.0750 \)

Degrees of freedom for the t tests for the coefficients = \( n - p - 1 = 12 \)
Because \( t > 0 \), \( p \)-value is twice the upper tail area for each
Using Excel: \( p \)-value for Intercept = \( 2*(1-\text{DIST}(3.0382,12,\text{TRUE})) = .0103 \)
Using Excel: \( p \)-value for X1 coefficient = \( 2*(1-\text{DIST}(3.6114,12,\text{TRUE})) = .0036 \)
Using Excel: \( p \)-value for X2 coefficient = \( 2*(1-\text{DIST}(5.0750,12,\text{TRUE})) = .0003 \)

\[b. \text{ The } p \text{-value (2 degrees of freedom numerator and 12 denominator) corresponding to } F = 71.92 \text{ is } .0000\]

Because the \( p \)-value \( \leq \alpha = .05 \), there is a significant relationship.

\[c. \text{ For } \beta_1: \text{ The } p \text{-value (12 degrees of freedom) corresponding to } t = 3.6114 \text{ is } .0036\]

Because the \( p \)-value \( \leq \alpha = .05 \), reject \( H_0: \beta_1 = 0 \)

For \( \beta_2 \): The \( p \)-value (12 degrees of freedom) corresponding to \( t = 5.0750 \) is .0003

Because the \( p \)-value \( \leq \alpha = .05 \), reject \( H_0: \beta_2 = 0 \)
48. a. The regression equation is

\[ \hat{y} = 14.4 - 8.69X_1 + 13.517X_2 \]

Standard error (top section) = 3.773 = \( s = \sqrt{\text{MSE}} \)

Therefore, \( \text{MSE} = s^2 = 3.773^2 = 14.2355 \)

Using algebra, we can use this to solve for \( n \) (number of observations):

\[
\frac{71.17}{14.2355} = 5 \text{ and } n - 3 = \frac{71.17}{14.2355} = 5 \text{ and therefore, } n = 8
\]

Regression df = \( p \) = number of independent variables = 2
Total df = \( n - 1 = 7 \)
Residual df = Total df – Regression df = 7 – 2 = 5
Regression df + Residual df = Total df

\[ \text{SSR} = \text{SST} - \text{SSE} = 720 - 71.17 = 648.83 \]

\[ R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{648.83}{720} = .90115 \]

\[ R_a^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} = 1 - (1 - .90115) \frac{8 - 1}{8 - 2 - 1} = .8616 \]

\[ \text{Multiple } R = r = \sqrt{R^2} \text{ with the same sign as the slope. Therefore, } r = \sqrt{.90115} = .9493 \]

\[ \text{MSR} = \frac{\text{SSR}}{p} = \frac{648.83}{2} = 324.415 \]

From earlier, \( s^2 = \text{MSE} = 14.2355 \)
\( F = \frac{MSR}{MSE} = \frac{324.415}{14.2355} = 22.79 \)

Using Excel, the \( p \)-value = \( F.DIST.RT(22.79,2,5) \) = .0031

\[ t = \frac{b}{s_b} \]

and therefore, for the Intercept, \( t = \frac{b}{s_b} = \frac{14.4}{8.191} = 1.7580 \)

For the X1 coefficient, \( t = \frac{b}{s_b} = \frac{-8.69}{1.555} = -5.5884 \)

For the X2 coefficient, \( t = \frac{b}{s_b} = \frac{13.517}{2.085} = 6.4830 \)

Degrees of freedom for the t tests for the coefficients = \( n - p - 1 = 5 \)

Because \( t < 0 \) for the X1 coefficient, \( p \)-value is two times the lower tail area

Using Excel: \( p \)-value for X1 coefficient = \( 2*(T.DIST(-5.5884,5,TRUE)) \) = .0025

For the Intercept and X2 Coefficient, \( t > 0 \), \( p \)-value is two times the upper tail area for each

Using Excel: \( p \)-value for Intercept = \( 2*(1-T.DIST(1.7580,5,TRUE)) \) = .1391

Using Excel: \( p \)-value for X2 coefficient = \( 2*(1-T.DIST(6.4830,5,TRUE)) \) = .0013

b. The \( p \)-value (2 degrees of freedom numerator and 5 degrees of freedom denominator) corresponding to \( F = 22.79 \) is .0031

Because the \( p \)-value is \( \leq \alpha = .05 \), there is a significant relationship.

c. \( R^2 = .9012 \)

\( R^2 = .8616 \)

good fit

d. For \( \beta_1 \): the \( p \)-value (5 degrees of freedom) corresponding to \( t = -5.5884 \) is .0025.

Because the \( p \)-value is \( \leq \alpha = .05 \), reject \( H_0: \beta_1 = 0 \)

For \( \beta_2 \): the \( p \)-value corresponding to \( t = 6.4830 \) is .0013

Because the \( p \)-value is \( \leq \alpha = .05 \), reject \( H_0: \beta_2 = 0 \)

50. a. Using datafile 2012FuelEcon and Excel’s Descriptive Statistics Regression Tool, the Regression tool output follows.

\[ \begin{array}{c}
\text{Regression Statistics} \\
\text{Multiple R} & 0.8013 \\
\text{R Square} & 0.6421 \\
\text{Adjusted R Square} & 0.6409 \\
\text{Standard Error} & 3.4123 \\
\text{Observations} & 309 \\
\end{array} \]

\[ \begin{array}{c}
\text{ANOVA} \\
\text{df} & SS & MS & F & \text{Significance F} \\
\end{array} \]

© 2018 Cengage Learning. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.
Multiple Regression

\[ \hat{y} = 41.0534 - 3.7232 \text{Displacement} \]

Because the \( p \)-value corresponding to \( F = 550.8029 \) is \( .0000 < \alpha = .05 \), there is a significant relationship.

b. Using datafile 2012FuelEcon and Excel’s Descriptive Statistics Regression Tool, the Excel Regression tool output follows.

\[
\begin{array}{cccc}
\text{Coefficients} & \text{Standard Error} & \text{t Stat} & \text{P-value} \\
\text{Intercept} & 41.0534 & 0.5166 & 79.4748 & 8.1E-207 \\
\text{Displacement} & -3.7232 & 0.1586 & -23.4692 & 1.8E-70 \\
\end{array}
\]

\[ \hat{y} = 40.5946 - 3.1944 \text{Displacement} - 2.7230 \text{FuelPremium} \]

c. For FuelPremium, the \( p \)-value corresponding to \( t = -6.4498 \) is \( .0000 < \alpha = .05 \); significant. The addition of the dummy variables is significant.

d. Using datafile 2012FuelEcon and Excel’s Descriptive Statistics Regression Tool, the Excel Regression tool output follows.

\[
\begin{array}{cccc}
\text{Coefficients} & \text{Standard Error} & \text{t Stat} & \text{P-value} \\
\text{Intercept} & 40.5946 & 0.4906 & 82.7379 & 1.8E-211 \\
\text{Displacement} & -3.1944 & 0.1701 & -18.7745 & 7.43E-53 \\
\text{FuelPremium} & -2.7230 & 0.4222 & -6.4498 & 4.37E-10 \\
\end{array}
\]
Observations 309

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>4</td>
<td>7308.5436</td>
<td>1827.1359</td>
<td>207.3108</td>
<td>1.5479E-85</td>
</tr>
<tr>
<td>Residual</td>
<td>304</td>
<td>2679.3075</td>
<td>8.8135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>308</td>
<td>9987.8511</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.7892</td>
<td>48.1055</td>
<td>3.5E-144</td>
</tr>
<tr>
<td>Displacement</td>
<td>0.1941</td>
<td>-16.7007</td>
<td>6.97E-45</td>
</tr>
<tr>
<td>FuelPremium</td>
<td>0.4519</td>
<td>-4.7253</td>
<td>3.52E-06</td>
</tr>
<tr>
<td>FrontWheel</td>
<td>0.5394</td>
<td>5.7005</td>
<td>2.83E-08</td>
</tr>
<tr>
<td>RearWheel</td>
<td>0.5413</td>
<td>6.1174</td>
<td>2.92E-09</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 37.9626 - 3.2418 \text{Displacement} - 2.1352 \text{FuelPremium} + 3.0747 \text{FrontWheel} + 3.3114 \text{RearWheel} \]

Since the p-value corresponding to \( F = 207.3108 \) is .0000 \( \leq \alpha = .05 \), there is a significant overall relationship. Because the p-values for each independent variable are also \( \leq \alpha = .05 \), each of the independent variables is significant.

52. a. Using datafileNBAStats and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

Regression Statistics

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.7339</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Square</td>
<td>0.5386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.5221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>10.7930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>30</td>
<td></td>
<td></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>3807.7298</td>
<td>3807.7298</td>
<td>32.6876</td>
<td>3.9263E-06</td>
</tr>
<tr>
<td>Residual</td>
<td>28</td>
<td>3261.6772</td>
<td>116.4885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>7069.407</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-294.7669</td>
<td>-4.8857</td>
<td>3.79099E-05</td>
</tr>
<tr>
<td>FG%</td>
<td>7.6966</td>
<td>5.7173</td>
<td>3.92633E-06</td>
</tr>
</tbody>
</table>

\[ \hat{y} = -294.7669 + 7.6966 \text{FG\%} \]

Since the p-value corresponding to \( t = 5.7173 \) or \( F = 32.6876 \) is .000 \( \leq \alpha = .05 \), there is a significant relationship between the percentage of games won and the percentage of field goals made.

b. An increase of 1% in the percentage of field goals made will increase the percentage of games won by approximately 7.7%.

e. Using datafileNBAStats and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:
### Regression Statistics

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.8764</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Square</td>
<td>0.7680</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.7197</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>8.2663</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>30</td>
<td></td>
<td></td>
<td></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>5</td>
<td>5429.4550</td>
<td>1085.8910</td>
<td>15.8916</td>
<td>6.13714E-07</td>
</tr>
<tr>
<td>Residual</td>
<td>24</td>
<td>1639.9520</td>
<td>68.3313</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>7069.407</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-407.9703</td>
<td>-5.9166</td>
<td>4.18419E-06</td>
</tr>
<tr>
<td>FG%</td>
<td>4.9612</td>
<td>3.6276</td>
<td>0.0013</td>
</tr>
<tr>
<td>3P%</td>
<td>2.3749</td>
<td>2.9413</td>
<td>0.0071</td>
</tr>
<tr>
<td>FT%</td>
<td>0.0049</td>
<td>0.0095</td>
<td>0.9925</td>
</tr>
<tr>
<td>RBOff</td>
<td>3.4612</td>
<td>2.5711</td>
<td>0.0168</td>
</tr>
<tr>
<td>RBDef</td>
<td>3.6853</td>
<td>2.8425</td>
<td>0.0090</td>
</tr>
</tbody>
</table>

\[
\hat{y} = -407.9703 + 4.9612 \text{ FG\%} + 2.3749 \text{ 3P\%} + 0.0049 \text{ FT\%} + 3.4612 \text{ RBOff} + 3.6853 \text{ RBDef}
\]

d. For the estimated regression equation developed in part (c), the percentage of free throws made (FT\%) is not significant because the p-value corresponding to \(t = .0095\) is \(0.9925 > \alpha = .05\). After removing this independent variable, using datafileNBAStats and Excel’s Descriptive Statistics Regression Tool, the Excel output is shown below:

### Regression Statistics

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.8764</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Square</td>
<td>0.7680</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.7309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>8.0993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>30</td>
<td></td>
<td></td>
<td></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>4</td>
<td>5429.4489</td>
<td>1357.3622</td>
<td>20.6920</td>
<td>1.24005E-07</td>
</tr>
<tr>
<td>Residual</td>
<td>25</td>
<td>1639.9581</td>
<td>65.5983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>7069.407</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-407.5790</td>
<td>-7.5178</td>
<td>7.1603E-08</td>
</tr>
<tr>
<td>FG%</td>
<td>4.9621</td>
<td>3.7125</td>
<td>0.0010</td>
</tr>
<tr>
<td>3P%</td>
<td>2.3736</td>
<td>3.0401</td>
<td>0.0055</td>
</tr>
<tr>
<td>RBOff</td>
<td>3.4579</td>
<td>2.7089</td>
<td>0.0120</td>
</tr>
<tr>
<td>RBDef</td>
<td>3.6859</td>
<td>2.9048</td>
<td>0.0076</td>
</tr>
</tbody>
</table>
\[ \hat{y} = -407.5790 + 4.9621 \text{FG\%} + 2.3736 \text{3P\%} + 3.4579 \text{RBOff} + 3.6859 \text{RDDef} \]

e. \[ \hat{y} = -407.5790 + 4.9621(45) + 2.3736(35) + 3.4579(12) + 3.6859(30) = 50.9\% \]