Chapter 1 Project

Americans, Here’s Looking at You

The U.S. Census Bureau annually publishes the Statistical Abstract of the United States, a 1000+ page book that provides us with a statistical insight into many of the most obscure and unusual facets of our lives. This is only one of thousands of sources for all kinds of things you have always wanted to know about and never thought to ask about. Are you interested in how many hours we work and play? How much we spend on snack foods? How the price of Red Delicious apples has gone up? All this and more—much more—can be found in the Statistical Abstract (http://www.census.gov/statab/www).

The statistical tidbits that follow come from a variety of sources and represent only a tiny sampling of what can be learned about Americans statistically. Take a look.

Communication Method Preferred
By Workers

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-mail</td>
<td>32%</td>
</tr>
<tr>
<td>Telephone</td>
<td>24%</td>
</tr>
<tr>
<td>Direct mail</td>
<td>18%</td>
</tr>
<tr>
<td>Personal letter</td>
<td>17%</td>
</tr>
</tbody>
</table>

Workers say they would rather be contacted by companies they do business with via e-mail than any other way.

Would You Like To See Your 100th Birthday?

- Yes: 63%
- No: 32%
- Don’t know: 5%

Should The Penny Be Eliminated?

- Yes: 23%
- No: 59%
- Not sure: 18%

Nearly 6 in 10 Americans want the penny to remain in circulation.

Driving Is Leading Danger To Teens

About 3500 teenagers died in teen-driven vehicles in the United States in 2003—a death toll that tops that of any disease or injury for teens.

<table>
<thead>
<tr>
<th>Driver age</th>
<th>Fatal crash involvement per 100 million miles traveled</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>9.3</td>
</tr>
<tr>
<td>17</td>
<td>8.3</td>
</tr>
<tr>
<td>18</td>
<td>6.5</td>
</tr>
<tr>
<td>19</td>
<td>7.2</td>
</tr>
<tr>
<td>20–24</td>
<td>4.3</td>
</tr>
<tr>
<td>25–29</td>
<td>2.3</td>
</tr>
<tr>
<td>30–39</td>
<td>1.6</td>
</tr>
<tr>
<td>40–69</td>
<td>1.6</td>
</tr>
<tr>
<td>70+</td>
<td>4.1</td>
</tr>
</tbody>
</table>

16-year-old drivers have highest fatal crash rate.
The chapter project is a way to assess what we have learned in this chapter. Study the statistical information presented by the graphs and charts, and ask yourself how the terms (population, sample, variable, statistic, type of variable) studied in this chapter apply to each and how you compare to the statistical story being told.

**Putting Chapter 1 to Work**

1. Write a 50-word paragraph describing what the word *statistics* means to you right now.

2. Write a 50-word paragraph describing what the word *random* means to you right now.

3. Write a 50-word paragraph describing what the word *sample* means to you right now.

2 With respect to the four graphics complete the following:

   a. What statistical population is of concern for all of these graphics?

   b. Identify one specific graphic. What variables were used to collect the information needed to determine the statistics reported?

   c. Name one statistic that is being reported in your graphic.

   d. To obtain the data for your graphic, what methods do you think were used: convenience sample, volunteer sample, random sample, survey, observational study, experiment, or judgment sample?

   e. Considering the method, how much faith do you have in the printed statistics? Describe possible biases.

**Your Study**

3 Select one of the “Americans, Here’s Looking at You” graphics then, using the students at your school or college as the population of concern, collect sample data from 30 students and produce your own version of the graphic. Write a paragraph describing how your results compare to those reported in the selected graphic.

4 Find an article or an advertisement in a newspaper or magazine that exemplifies the use of statistics.

   a. Identify and describe one statistic reported in the article.

   b. Identify and describe the variable related to the statistic in part a.

   c. Identify and describe the sample related to the statistic in part a.

   d. Identify and describe the population from which the sample in part c was taken.
You and the Internet

Ever wonder what other people do when they are on the Internet? Well, you are not the only one. Stanford Institute for the Quantitative Study of Society (SIQSS) supported a study that looked at how people use the Internet. Four thousand respondents were asked to select which of 17 common Internet activities they did or did not do. E-mail was identified by 90% of the respondents as one of their many uses of the Internet. Other common uses included information gathering, entertainment, chat rooms, and business transactions.

The preceding graph summarizes all the information from the study of 4000 Internet users. Can you imagine if all of that information was written out in sentences? Graphical displays (pictures) can truly be worth a thousand words. Not only is the information in a clearer, more concise format, but the format also allows us to make some conclusions at the same time. We immediately know which activities most users engage in and which activities are not that popular.

If you had been asked, “What activity do you do the most on the Internet?” how would you have answered? Do you believe your answer is represented accurately in the diagram?

“You and the Internet” is a way to assess what we have learned in this chapter. Based on the percents stated in the “What Users Do on the Internet” graph, respondents were able to pick more than one Internet activity. If you were asked, how many of the listed activities would you select as something you do? Suppose a sample of students was asked about their Internet activities. Would their answers differ from yours? Would they differ from the 4000 respondents? “Putting Chapter 2 to Work” will help us answer these questions.

Putting Chapter 2 to Work

1 a. How many different Internet activities did you engage in last week?
b. How do you think you compare to all Internet users?

2 [CP02] Students in a statistics course offered over the Internet were asked how many different Internet activities they engaged in during a typical week. The following data show the number of activities:

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

a. List all types of charts and graphs shown in Chapter 2 that would be appropriate for use with the set of 40 data listed.

b. What types of graphs would not be appropriate? Explain why.

c. Display the data using each of the charts and graphs listed in part a.

d. Which graph do you think best represents the data? Explain why.

e. Find the five measures of central tendency for these data (mean, median, mode, midrange, and midquartile).

f. Find the three measures of dispersion for the data (range, variance, and standard deviation).

g. Find the value of several measures of position: $P_{10}$, $Q_1$, $Q_3$, $P_{90}$, and $P_{95}$.

h. How many different Internet activities do you engage in during a typical week? Using the mean and standard deviation calculated in parts e and f, determine your $z$-score. What is this telling you about yourself with respect to statistics students’ Internet usage?

i. Use one graph from part c plus at least one measure of central tendency and one measure of dispersion, and write a description of statistics students’ Internet usage, number of Internet activities per week.

j. According to the empirical rule, if the distribution is normal, approximately 68% of the number of different Internet activities engaged in by statistics students will fall between what two values? Is this true? Why or why not?

k. According to Chebyshev’s theorem, approximately 75% of the number of different Internet activities engaged in by statistics students will fall within what two values? Is this true? Why or why not?

l. The sample information pictured in the “What Users Do on the Internet” graph is different than, but related to, the sample information you have been working with in parts a through k. Describe the data collected for the graph in “What Users Do on the Internet” and explain how they differ from the data listed here.

Your Study

3 a. Design your own study of Internet usage. Define a specific population that you will sample, describe your sampling plan, collect your data, and answer parts c through l in “Putting Chapter 2 to Work,” question 2.

b. Discuss the differences and similarities between the Internet usage described by the sample of 40 statistics students (given in question 2) and your sample.
**Chapter 3 Project**

**The Kid Is All Grown Up**

MINNEAPOLIS—The Kid is all grown up, and he has an NBA MVP award to prove it. Kevin Garnett got 120 of 123 first-place votes to beat two-time winner Tim Duncan for the honor Monday, three days after his Minnesota Timberwolves won a playoff series for the first time. Garnett’s teammates attended a packed news conference at the Timberwolves’ arena, and he praised them repeatedly.

Playing everywhere from center to point guard, the 7-footer averaged 24.2 points, a league-leading 13.9 rebounds and 5.0 assists this season—and his playoff stats are even better. Garnett joined Larry Bird as the only players to average 20 points, 10 rebounds and five assists for five consecutive years.

Nicknamed “The Kid,” Garnett made the All-Star team in his second season, and his success helped fuel the wave of preps-to-pros players.

GARNETT EARNS 120 OF 123 FIRST-PLACE VOTES
Associated Press
Monday, May 4, 2004

<table>
<thead>
<tr>
<th>Player</th>
<th>Personal Fouls per Game</th>
<th>Points per Game</th>
<th>Player</th>
<th>Personal Fouls per Game</th>
<th>Points per Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garnett</td>
<td>2.5</td>
<td>24.2</td>
<td>Madsen</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td>Cassell</td>
<td>3</td>
<td>19.8</td>
<td>Martin</td>
<td>1.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Spiewell</td>
<td>1.2</td>
<td>16.8</td>
<td>McLeod</td>
<td>1.2</td>
<td>2.7</td>
</tr>
<tr>
<td>Szczerbiak</td>
<td>1.5</td>
<td>10.2</td>
<td>Goldwire</td>
<td>1.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Hudson</td>
<td>1.1</td>
<td>7.5</td>
<td>Miller</td>
<td>1.9</td>
<td>2.5</td>
</tr>
<tr>
<td>Hoiberg</td>
<td>1.7</td>
<td>6.7</td>
<td>Johnson</td>
<td>2.4</td>
<td>1.9</td>
</tr>
<tr>
<td>Olowokandi</td>
<td>3.2</td>
<td>6.5</td>
<td>Lewis</td>
<td>0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Trett</td>
<td>1.9</td>
<td>5.6</td>
<td>Ebi</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Hassell</td>
<td>2.5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: http://sports.espn.go.com/nba/teams

Do you play basketball or at least follow the sport? Does it seem like those that make more baskets also have the most personal fouls? Think of those players who do not score any points, do they still make personal fouls? Now if you do not play basketball or even follow it, you still know about relationships. Just think about yourself. Does it seem that as you grew taller, your shoe size was also increasing? Is there a relationship between a person’s height and his or her shoe size? Does it seem that those students who study more get better grades? Is there a relationship between hours studied and grades? Does it seem that those students who travel farther one way to school also need more time to travel to school?

As a way of assessing the statistical techniques for bivariate data that we have learned in this chapter, let’s return to “The Kid Is All Grown Up.” For any basketball player, the number of points scored per game and the number of personal fouls committed per game are of interest. Could a clear and definite relationship exist between these two variables, and if so, why?

**Putting Chapter 3 to Work**

1. Is there a relationship (pattern) between the two variables, points scored per game and number of personal fouls committed per game? Explain why or why not.
2. Do you think it is reasonable (or possible) to predict the number of points scored based on the
Chapter 3 Project

number of personal fouls committed per game for a Minnesota Timberwolves’ player? Explain why or why not.

2 a. Construct a scatter diagram, using points scored per game, y, and number of personal fouls committed per game, x. Explain why you believe there is or is not a relationship.

b. Are the two variables points scored per game and number of personal fouls committed per game correlated? Use the correlation coefficient to justify your answer.

c. Express the relationship between the two variables total points scored, y, and number of personal fouls committed, x, as a linear equation.

d. Using the results from part c, if a Minnesota Timberwolves player committed two fouls in a game, how many points would you expect him to score?

e. If the player in part d committed a third personal foul, how many extra points would you expect him to score?

f. How does the slope for the line of best fit relate to the number of additional points expected when the player commits one extra personal foul?

g. Do the preceding results show a cause-and-effect relationship between total points scored and number of personal fouls committed? Explain.

h. Should the team coach instruct a player to commit an extra personal foul so that he will score more points? Explain.

i. Name at least one possible lurking variable for the preceding situation.

Suppose the preceding investigation is to be expanded to include an additional variable, “minutes played per game.”

j. Describe the relationship you think might exist between the variables “minutes played per game” and “points scored per game.” Explain why.

k. Describe the relationship you think might exist between the variables “minutes played per game” and “number of personal fouls committed per game.” Explain why.

l. Could “minutes played per game” be a lurking variable for the work completed in parts a–h? Explain.

Your Study

3 a. The situation described in question 2 only occurred with the Minnesota Timberwolves during the 2003–2004 regular season. Use the Internet (search by the team name) to obtain the season team statistics for your favorite professional or intercollegiate basketball team, or see the coach of a local high school or college team.

b. Answer the same questions asked in question 2 for your selected team.

c. Discuss the differences and similarities between the Minnesota Timberwolves and your selected team. Consider other lurking variables.
Part 1 Project

Working with Your Own Data

Each semester, new students enter your college environment. You may have wondered, What will the student body be like this semester? As a beginning statistics student, you have just finished studying three chapters of basic descriptive statistical techniques. You can use some of these techniques to describe some characteristics of your college’s student body.

A Single-Variable Data
1. Define the population to be studied.
2. Choose a variable to define. (You may define your own variable, or you may use one of the variables in the accompanying table* if you are not able to collect your own data. Ask your instructor for guidance.)
3. Collect 35 pieces of data for your variable.
4. Construct a stem-and-leaf display of your data. (Be sure to label it.)
5. Calculate the value of the measure of central tendency that you believe best answers the question: What is the average value of your variable? Explain why you chose this measure.
6. Calculate the sample mean for your data (unless you used the mean in question 5).
7. Calculate the sample standard deviation for your data.
8. Find the value of the 85th percentile, \( P_{85} \).
9. Construct a graphic display (other than a stem-and-leaf) that you believe “best” displays your data. Explain why the graph best presents your data.
10. Write a summary paragraph describing your findings.

B Bivariate Data
1. Define the population to be studied.
2. Choose and define two quantitative variables that will produce bivariate data. (You may define your own variables, or you may use two of the variables in the accompanying table if you are not able to collect your own data. Ask your instructor for guidance.)
3. Collect 15 ordered pairs of data.
4. Construct a scatter diagram of your data. (Be sure to label it completely.)
5. Using a table to assist with the organization, calculate the extensions \( x^2 \), \( xy \), and \( y^2 \), and the summations of \( x \), \( y \), \( x^2 \), \( xy \), and \( y^2 \).
6. Calculate the linear correlation coefficient, \( r \).
7. Calculate the equation of the line of best fit.
8. Draw the line of best fit on your scatter diagram.
9. Write a summary paragraph describing your findings.

*The table of data on the next page was collected on the first day of class last semester. You may use it as a source for your data if you are not able to collect your own.

Variable A: student’s gender (male/female)
Variable B: student’s age at last birthday
Variable C: number of completed credit hours toward degree
Variable D: “Do you have a job (full/part time)?” (yes/no)
Variable E: number of hours worked last week, if \( D = \text{yes} \)
Variable F: wages (before taxes) earned last week, if \( D = \text{yes} \)

FYI The computer will select your random sample.
### Part 1 Project

#### [DS1]

<table>
<thead>
<tr>
<th>Student</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Student</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>21</td>
<td>16</td>
<td>No</td>
<td></td>
<td></td>
<td>51</td>
<td>F</td>
<td>42</td>
<td>34</td>
<td>Yes</td>
<td>40</td>
<td>244</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>18</td>
<td>0</td>
<td>Yes</td>
<td>10</td>
<td>34</td>
<td>52</td>
<td>M</td>
<td>25</td>
<td>60</td>
<td>Yes</td>
<td>60</td>
<td>503</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>23</td>
<td>18</td>
<td>Yes</td>
<td>46</td>
<td>206</td>
<td>53</td>
<td>M</td>
<td>39</td>
<td>32</td>
<td>Yes</td>
<td>40</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>17</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
<td>54</td>
<td>M</td>
<td>29</td>
<td>13</td>
<td>Yes</td>
<td>39</td>
<td>375</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>17</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>157</td>
<td>55</td>
<td>M</td>
<td>19</td>
<td>18</td>
<td>Yes</td>
<td>51</td>
<td>201</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>40</td>
<td>17</td>
<td>No</td>
<td></td>
<td></td>
<td>56</td>
<td>M</td>
<td>25</td>
<td>0</td>
<td>Yes</td>
<td>48</td>
<td>500</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>20</td>
<td>16</td>
<td>Yes</td>
<td>40</td>
<td>300</td>
<td>57</td>
<td>F</td>
<td>18</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>18</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
<td>58</td>
<td>M</td>
<td>32</td>
<td>68</td>
<td>Yes</td>
<td>44</td>
<td>473</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>18</td>
<td>0</td>
<td>Yes</td>
<td>20</td>
<td>70</td>
<td>59</td>
<td>F</td>
<td>21</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>29</td>
<td>9</td>
<td>Yes</td>
<td>8</td>
<td>32</td>
<td>60</td>
<td>F</td>
<td>26</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>320</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>20</td>
<td>22</td>
<td>Yes</td>
<td>38</td>
<td>146</td>
<td>61</td>
<td>M</td>
<td>24</td>
<td>11</td>
<td>Yes</td>
<td>45</td>
<td>330</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>34</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>340</td>
<td>62</td>
<td>F</td>
<td>19</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>220</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>19</td>
<td>31</td>
<td>Yes</td>
<td>29</td>
<td>105</td>
<td>63</td>
<td>M</td>
<td>19</td>
<td>0</td>
<td>Yes</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>18</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
<td>64</td>
<td>F</td>
<td>35</td>
<td>59</td>
<td>Yes</td>
<td>25</td>
<td>88</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>20</td>
<td>0</td>
<td>Yes</td>
<td>48</td>
<td>350</td>
<td>65</td>
<td>F</td>
<td>24</td>
<td>6</td>
<td>Yes</td>
<td>40</td>
<td>300</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>27</td>
<td>3</td>
<td>Yes</td>
<td>40</td>
<td>130</td>
<td>66</td>
<td>F</td>
<td>20</td>
<td>33</td>
<td>Yes</td>
<td>40</td>
<td>170</td>
</tr>
<tr>
<td>17</td>
<td>M</td>
<td>19</td>
<td>10</td>
<td>Yes</td>
<td>40</td>
<td>202</td>
<td>67</td>
<td>F</td>
<td>26</td>
<td>0</td>
<td>Yes</td>
<td>52</td>
<td>300</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>18</td>
<td>16</td>
<td>Yes</td>
<td>40</td>
<td>140</td>
<td>68</td>
<td>F</td>
<td>17</td>
<td>0</td>
<td>Yes</td>
<td>27</td>
<td>100</td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>19</td>
<td>4</td>
<td>Yes</td>
<td>6</td>
<td>22</td>
<td>69</td>
<td>M</td>
<td>25</td>
<td>18</td>
<td>Yes</td>
<td>41</td>
<td>355</td>
</tr>
<tr>
<td>20</td>
<td>F</td>
<td>29</td>
<td>9</td>
<td>No</td>
<td></td>
<td></td>
<td>70</td>
<td>M</td>
<td>24</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>21</td>
<td>0</td>
<td>Yes</td>
<td>20</td>
<td>80</td>
<td>71</td>
<td>M</td>
<td>21</td>
<td>0</td>
<td>Yes</td>
<td>30</td>
<td>150</td>
</tr>
<tr>
<td>22</td>
<td>F</td>
<td>39</td>
<td>6</td>
<td>Yes</td>
<td>8</td>
<td>32</td>
<td>72</td>
<td>M</td>
<td>30</td>
<td>12</td>
<td>Yes</td>
<td>48</td>
<td>555</td>
</tr>
<tr>
<td>23</td>
<td>M</td>
<td>23</td>
<td>34</td>
<td>Yes</td>
<td>42</td>
<td>415</td>
<td>73</td>
<td>F</td>
<td>19</td>
<td>0</td>
<td>Yes</td>
<td>38</td>
<td>169</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td>31</td>
<td>0</td>
<td>Yes</td>
<td>48</td>
<td>325</td>
<td>74</td>
<td>M</td>
<td>32</td>
<td>45</td>
<td>Yes</td>
<td>40</td>
<td>385</td>
</tr>
<tr>
<td>25</td>
<td>F</td>
<td>22</td>
<td>7</td>
<td>Yes</td>
<td>40</td>
<td>195</td>
<td>75</td>
<td>M</td>
<td>26</td>
<td>90</td>
<td>Yes</td>
<td>40</td>
<td>340</td>
</tr>
<tr>
<td>26</td>
<td>F</td>
<td>27</td>
<td>75</td>
<td>Yes</td>
<td>20</td>
<td>130</td>
<td>76</td>
<td>M</td>
<td>20</td>
<td>64</td>
<td>Yes</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>27</td>
<td>F</td>
<td>19</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
<td>77</td>
<td>M</td>
<td>24</td>
<td>0</td>
<td>Yes</td>
<td>30</td>
<td>150</td>
</tr>
<tr>
<td>28</td>
<td>M</td>
<td>22</td>
<td>20</td>
<td>Yes</td>
<td>40</td>
<td>470</td>
<td>78</td>
<td>M</td>
<td>20</td>
<td>14</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>60</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>390</td>
<td>79</td>
<td>M</td>
<td>21</td>
<td>70</td>
<td>Yes</td>
<td>40</td>
<td>340</td>
</tr>
<tr>
<td>30</td>
<td>M</td>
<td>25</td>
<td>14</td>
<td>No</td>
<td></td>
<td></td>
<td>80</td>
<td>F</td>
<td>20</td>
<td>13</td>
<td>Yes</td>
<td>40</td>
<td>206</td>
</tr>
<tr>
<td>31</td>
<td>F</td>
<td>24</td>
<td>45</td>
<td>No</td>
<td></td>
<td></td>
<td>81</td>
<td>F</td>
<td>33</td>
<td>3</td>
<td>Yes</td>
<td>32</td>
<td>246</td>
</tr>
<tr>
<td>32</td>
<td>M</td>
<td>34</td>
<td>4</td>
<td>No</td>
<td></td>
<td></td>
<td>82</td>
<td>F</td>
<td>25</td>
<td>68</td>
<td>Yes</td>
<td>40</td>
<td>330</td>
</tr>
<tr>
<td>33</td>
<td>M</td>
<td>29</td>
<td>48</td>
<td>No</td>
<td></td>
<td></td>
<td>83</td>
<td>F</td>
<td>29</td>
<td>48</td>
<td>Yes</td>
<td>40</td>
<td>525</td>
</tr>
<tr>
<td>34</td>
<td>M</td>
<td>22</td>
<td>80</td>
<td>Yes</td>
<td>40</td>
<td>336</td>
<td>84</td>
<td>F</td>
<td>40</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>400</td>
</tr>
<tr>
<td>35</td>
<td>M</td>
<td>21</td>
<td>12</td>
<td>Yes</td>
<td>26</td>
<td>143</td>
<td>85</td>
<td>F</td>
<td>36</td>
<td>3</td>
<td>Yes</td>
<td>40</td>
<td>300</td>
</tr>
<tr>
<td>36</td>
<td>F</td>
<td>18</td>
<td>0</td>
<td>No</td>
<td></td>
<td></td>
<td>86</td>
<td>F</td>
<td>35</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>280</td>
</tr>
<tr>
<td>37</td>
<td>M</td>
<td>18</td>
<td>0</td>
<td>Yes</td>
<td>13</td>
<td>65</td>
<td>87</td>
<td>F</td>
<td>28</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>350</td>
</tr>
<tr>
<td>38</td>
<td>M</td>
<td>40</td>
<td>64</td>
<td>Yes</td>
<td>40</td>
<td>390</td>
<td>88</td>
<td>F</td>
<td>27</td>
<td>9</td>
<td>Yes</td>
<td>40</td>
<td>260</td>
</tr>
<tr>
<td>39</td>
<td>F</td>
<td>31</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>200</td>
<td>89</td>
<td>F</td>
<td>26</td>
<td>3</td>
<td>Yes</td>
<td>40</td>
<td>240</td>
</tr>
<tr>
<td>40</td>
<td>F</td>
<td>32</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>270</td>
<td>90</td>
<td>F</td>
<td>23</td>
<td>9</td>
<td>Yes</td>
<td>40</td>
<td>330</td>
</tr>
<tr>
<td>41</td>
<td>F</td>
<td>37</td>
<td>0</td>
<td>Yes</td>
<td>24</td>
<td>150</td>
<td>91</td>
<td>M</td>
<td>41</td>
<td>3</td>
<td>Yes</td>
<td>23</td>
<td>253</td>
</tr>
<tr>
<td>42</td>
<td>F</td>
<td>35</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>350</td>
<td>92</td>
<td>M</td>
<td>39</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>110</td>
</tr>
<tr>
<td>43</td>
<td>M</td>
<td>21</td>
<td>72</td>
<td>Yes</td>
<td>45</td>
<td>470</td>
<td>93</td>
<td>M</td>
<td>21</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>246</td>
</tr>
<tr>
<td>44</td>
<td>F</td>
<td>27</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>550</td>
<td>94</td>
<td>F</td>
<td>32</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>350</td>
</tr>
<tr>
<td>45</td>
<td>F</td>
<td>42</td>
<td>47</td>
<td>Yes</td>
<td>37</td>
<td>300</td>
<td>95</td>
<td>F</td>
<td>48</td>
<td>58</td>
<td>Yes</td>
<td>40</td>
<td>714</td>
</tr>
<tr>
<td>46</td>
<td>F</td>
<td>41</td>
<td>21</td>
<td>Yes</td>
<td>40</td>
<td>250</td>
<td>96</td>
<td>F</td>
<td>26</td>
<td>0</td>
<td>Yes</td>
<td>32</td>
<td>200</td>
</tr>
<tr>
<td>47</td>
<td>M</td>
<td>36</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>400</td>
<td>97</td>
<td>F</td>
<td>27</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>350</td>
</tr>
<tr>
<td>48</td>
<td>M</td>
<td>25</td>
<td>16</td>
<td>Yes</td>
<td>40</td>
<td>480</td>
<td>98</td>
<td>F</td>
<td>52</td>
<td>56</td>
<td>Yes</td>
<td>40</td>
<td>390</td>
</tr>
<tr>
<td>49</td>
<td>F</td>
<td>18</td>
<td>0</td>
<td>Yes</td>
<td>45</td>
<td>189</td>
<td>99</td>
<td>F</td>
<td>34</td>
<td>27</td>
<td>Yes</td>
<td>8</td>
<td>77</td>
</tr>
<tr>
<td>50</td>
<td>M</td>
<td>22</td>
<td>0</td>
<td>Yes</td>
<td>40</td>
<td>385</td>
<td>100</td>
<td>F</td>
<td>49</td>
<td>3</td>
<td>Yes</td>
<td>24</td>
<td>260</td>
</tr>
</tbody>
</table>
Chapter 4 Project

Sweet Statistics

Did you know that the idea for “M&M’s” Plain Chocolate Candies was born in the backdrop of the Spanish Civil War? Legend has it that on a trip to Spain, Forrest Mars Sr. encountered soldiers who were eating pellets of chocolate that were encased in a hard sugary coating to prevent them from melting. Mr. Mars was inspired by this concept and returned home and invented the recipe for “M&M’s” Plain Chocolate Candies.

Statistics class had started and the teacher was discussing percentages, proportions, and probabilities—how the three are similar and yet different. All of a sudden a student mentioned that she heard that the previous semester’s class did a lesson using, and eating, M&M’s; she asked if this year’s class would be doing something similar. The conversation was soon focused entirely on M&M’s—their color combinations and the percentage of each color. The 24 class members were each asked to guess the percentage of each color they believed were contained in those little brown bags of M&M’s Plain Chocolate Candies. They were told there would be a prize for the person whose guess was the closest to the actual number. Each student wrote down the percentages and turned them in; in return, the students received a little brown bag. “Oh, this is that lesson!” “Yes!” the teacher replied, “and before you open those bags, we must have a plan.” Each student was to count the number of M&M’s of each color in his or her bag and record the six counts; then the class totals would be determined. The resulting distribution of counts is shown in Table 1.

The class totals were converted to percentages (Table 2), and each student was asked to determine the six percentages they observed for their own bag of M&M’s.

The discussion that followed centered on the variation that occurred from bag to bag, with some students being quite surprised to see so much variation. Several bags either had none or only one of a color, and a few bags had a fairly large proportion of just one or two colors. Have you ever noticed either of these extremes when you opened a bag of M&M’s?

The percents reported in Table 2 are the percentages for each color found in this sample of 692 M&M’s. Percentages behave very much like probability numbers, but the question being asked in probability is quite different. In the preceding illustration, we are treating the information as sample data and describing the results found. If we now think in terms of a probability, we will turn the orientation around and treat the complete set of 692 M&M’s as the complete list of possibilities and ask questions about the likeliness of certain events when one M&M is randomly selected from the entire collection of 692.

For example, suppose we were to dump all 692 M&M’s into a large bowl and thoroughly mix them. Now consider the question, “If one M&M is selected at random from the bowl, what is the probability that it will be orange?” We hope that your thinking is along the following lines: selected randomly means each M&M has the same chance of being selected, and because there are 137 orange M&M’s in the bowl, the probability of selecting an orange M&M is 137/692, or 0.198.

You have seen the number 0.198 before, only it was expressed as 19.8%. Percentages and probability numbers are “the same, but different.” (You have probably heard that before, somewhere.) The numbers have the same value and behave with the same properties; however, the orientation of the situation and the questions asked are different, as you will see.

<table>
<thead>
<tr>
<th>Color</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>91</td>
</tr>
<tr>
<td>Yellow</td>
<td>112</td>
</tr>
<tr>
<td>Red</td>
<td>102</td>
</tr>
<tr>
<td>Blue</td>
<td>151</td>
</tr>
<tr>
<td>Orange</td>
<td>137</td>
</tr>
<tr>
<td>Green</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>692</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Color</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>13.2</td>
</tr>
<tr>
<td>Yellow</td>
<td>16.2</td>
</tr>
<tr>
<td>Red</td>
<td>14.7</td>
</tr>
<tr>
<td>Blue</td>
<td>21.8</td>
</tr>
<tr>
<td>Orange</td>
<td>19.8</td>
</tr>
<tr>
<td>Green</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>

TABLE 1
M&M Colors by Count

TABLE 1.2
M&M Colors by Percentages
Chapter 4 Project

The chapter project is a way to assess what we have learned in this chapter. And what a better way to do that than with some candy! We can explore the differences between theoretical and experimental probabilities as well as see the law of large numbers in action—all with M&M’s. Now that is “Sweet Statistics.” Let’s begin.

Putting Chapter 4 to Work

1. If you bought a bag of M&M’s, what color M&M would you expect to see the most? What color the least? Why?
   a. If you bought a bag of M&M’s, would you expect to find the percentages listed previously in Table 2? If not, why and what would you expect?

2. Let’s take a theoretical look at the expected. Mars, Inc., currently uses the following percentages to mix the colors for M&M’s Milk Chocolate Candies: 13% brown, 13% red, 14% yellow, 16% green, 20% orange, 24% blue.
   a. Construct a bar graph showing the expected (theoretical) proportion of M&M’s for each color.
   b. Theoretically, what percentage of red M&M’s should you expect in a bag of M&M’s?
   c. If you opened a bag of M&M’s right now, would you be surprised to find color percentages different from those given by Mars? Explain.

An empirical (experimental) look at what happened.

2. a. Obtain a pack of M&M’s (at least a 1.69 oz. size—approximately $0.50 in cost).
   b. Record the number of each color in a frequency distribution with the headings “Color” and “Frequency.”
   c. Verify the total number of M&M’s with the sum of the Frequency column.
   d. Now you may snack!

Your Study

3. a. Use a computer (or random-number table) to generate a random sample of 56 M&M’s, using the corresponding theoretical probabilities for each color.
   b. Form a frequency distribution of the random data.
   c. Construct a bar graph showing the relative frequencies for each color. Use the same color order as in part a of question 2.
   d. Compare your experimental findings with the theoretical expectations.
   e. Repeat parts a–d three more times.
   f. Describe the variability you observe between the samples.

An empirical (experimental) look at what happened.

3. a. Use a computer (or random-number table) to generate a random sample of 56 M&M’s, using the corresponding theoretical probabilities for each color.
   b. Form a frequency distribution of the random data.
   c. Construct a bar graph showing the relative frequencies for each color. Use the same color order as in part a of question 2.
   d. Compare your experimental findings with the theoretical expectations.
   e. Repeat parts a–d three more times.
   f. Describe the variability you observe between the samples.
   g. Consolidate your four frequency distributions into one frequency distribution having a frequency total of 224 M&M’s.
   h. Construct a bar graph of the consolidation showing relative frequencies for each color. Use the same color order as in part a of question 2.
   i. Compare these experimental findings with the theoretical expectations.
   j. Compare the consolidated findings with the four previous individual findings.
   k. How does the law of large numbers impact this mini study?

MINITAB and Excel can only generate random numbers. Therefore, it is common practice to use numbers in place of the colors (words). Use the numbers 1, 2, 3, 4, 5, 6 to correspond to brown, red, . . . , blue, respectively.
MINITAB

a. Input the numbers 1-6 into C1 and their corresponding probabilities in C2; then continue with:

Choose: Calc > Random Data > Discrete
Enter: Number of rows of data to generate: 56 (# of M&M's in a pack)
       Store in column(s): C3
       Values in: C1 (color numbers)
       Probabilities in: C2 > OK

b. To obtain the frequency distribution, continue with:

Choose: Stat > Tables > Cross Tabulation & Chi Square
Enter: Categorical variables: For rows: C3
Select: Display: Counts and Column percents > OK

c. To construct a bar graph enter the actual colors in C4 and the corresponding probabilities (%) found in step b in C5:

Choose: Graph > Bar Chart > Bar represent:
       Values from a table > One Column of values: Simple > OK
Enter: Graph variables: C5
       Categorical variables: C4
Select: Labels > Data Labels > Label Type: Use y-value labels > OK
Select: Data View > Data Display: Bars > OK > OK

Excel

a. Input the numbers 1-6 in column A and their corresponding probabilities in column B; then continue with:

Choose: Data > Data Analysis > Random Number Generation > OK
Enter: Number of Variables: 1
       Number of Random Numbers: 56 (# of M&M's in a pack)
       Distribution: Discrete
       Value & Prob. Input Range: (A1:B7 select data cells)
Select: Output range
Enter: (C1 or select cell) > OK

b. The frequency distribution is given with the histogram of the generated data. Use the histogram Excel commands on pages 53–54 using the data in column C and the bin range in column A.

c. Divide the frequencies by 56 to obtain the corresponding probabilities. Enter the actual colors in column D (ex. D13:D18) and the corresponding probabilities in column E (ex. E13:E18). To construct a bar graph, continue with:

Choose: Insert > Column > 1st picture(usually)
Enter: Chart and axes titles > Finish (Edit as needed)
Chapter 5 Project

Caffeine Drinking

Are Starbucks and other coffee purveyors taking over the country? It appears that way. One of the most common sights is a person on a cell phone with a cup of coffee. Think about it. How many people fitting that description have you seen today? Maybe you are one of them!

Consider the graphic “Americans Like Their Coffee!” It displays the number of cups or cans of caffeinated beverages adult Americans say they drink daily. The number of daily cups ranges from zero cups to four or more cups. Can you find yourself on the graphic?

Americans Like Their Coffee!

<table>
<thead>
<tr>
<th>Number of cups or cans per day</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>22%</td>
</tr>
<tr>
<td>One</td>
<td>16%</td>
</tr>
<tr>
<td>Two</td>
<td>21%</td>
</tr>
<tr>
<td>Three</td>
<td>16%</td>
</tr>
<tr>
<td>Four</td>
<td>25%</td>
</tr>
</tbody>
</table>

Source: Data from Shannon Reilly and Alejandro Gonzalez, © 2005 USA Today.

Who else might be interested in this information besides Starbucks? Apparently, the National Sleep Foundation is. Its mission and goals statement, as noted on its web page, is:

The National Sleep Foundation (NSF) is an independent nonprofit organization dedicated to improving public health and safety by achieving understanding of sleep and sleep disorders, and by supporting education, sleep-related research, and advocacy.

Source: [http://www.sleepfoundation.org](http://www.sleepfoundation.org)

Based on the poll of 1506 adults and the common belief that caffeine consumption affects sleep, does it appear that caffeine should be a concern of NSF? Probably not.
Let’s take a second look at “Caffeine Drinking” and test our knowledge of the material presented in this chapter. Based on the information from USA Today, we have the number of cups or cans of caffeinated beverages adult Americans say they drink daily and their corresponding probabilities. Consider where you might fit into this situation.

Putting Chapter 5 to Work

1 a. What variable could be used to describe all five of the events shown in the “Americans Like Their Coffee!” graphic?

b. Is the variable in part a discrete or continuous? Why?

c. Are events $x = 0, 1, 2, \ldots$ mutually exclusive? Explain why or why not.

d. What characteristics of a circle graph make it appropriate for use with a probability distribution? Be specific.

e. Construct a circle graph depicting the information described in the graphic.

f. Express the information in the circle graph as a probability distribution.

g. Assuming the information on the circle graph represents the population, find the mean and standard deviation of the variable described in part a.

h. Draw a histogram to display the information in the graphic. Describe the histogram. Is it a normal distribution? Explain.

i. Locate the mean and standard deviation found in part g on the histogram drawn in part h.

j. Do the empirical and Chebyshev’s rules apply? Justify your answer.

Your Study

2 Design your own study of caffeine drinking.

a. Define a specific population that you will sample, describe your sampling plan, and collect your data.

b. Express your sample as a relative frequency distribution and draw a histogram.

c. Express your sample as a frequency distribution and find the sample mean and sample standard deviation.

d. Discuss the differences and similarities between your sample and the distribution shown in the graphic “Americans Like Their Coffee!”
There are many kinds of aptitude tests. Some are for specific purposes, such as measurement of finger dexterity, something that might be important on a particular job. Others are of more general aptitudes. So-called intelligence tests are examples of general aptitude tests.

The Binet Intelligence Scale. Alfred Binet, who devised the first general aptitude test at the beginning of the 20th century, defined intelligence as the ability to make adaptations. The general purpose of the test was to determine which children in Paris could benefit from school. Binet’s test, like its subsequent revisions, consists of a series of progressively more difficult tasks that children of different ages can successfully complete. A child who can solve problems typically solved by children at a particular age level is said to have that mental age. For example, if a child can successfully do the same tasks that an average eight-year-old can do, he or she is said to have a mental age of eight. The intelligence quotient, or IQ, is defined by the formula:

\[
\text{intelligence quotient} = \frac{\text{mental age}}{\text{chronological age}} \times 100
\]

There has been a great deal of controversy in recent years over what intelligence tests measure. Many of the test items depend on either language or other specific cultural experiences for correct answers. Nevertheless, such tests can rather effectively predict school success. If school requires language and the tests measure language ability at a particular point of time in a child’s life, then the test is a better-than-chance predictor of school performance.

Deviation IQ Scores. Present-day tests of intelligence or other abilities use deviation scores. These scores represent the deviation of a particular person from the average score for similar persons. Suppose you take a “general aptitude test” and get a score of 115. This does not mean that your mental age is greater than your chronological age; it means that you are “above average” in some degree. Because we have become accustomed to thinking of an IQ score of 100 as average, most general aptitude tests are scored in such a way that 100 is average. A person scoring 115 would generally have a score higher than the scores of about 85 percent of people who take the test; a score of 84 would be better than about 16 percent. The exact interpretation of a test score depends on the particular test, but Figure 2.2 shows how the scores on a number of commonly used aptitude tests are interpreted in terms of how an individual compares with a group.

Figure 2.2 pictures the comparison of several deviation scores and the nor-
Chapter 6 Project

All normal probability distributions have the same shape and distribution relative to the mean and standard deviation. In this chapter we learned how to use the standard normal probability distribution to answer questions about all normal distributions. Let’s return to distribution of IQ scores discussed in “Intelligence Scores” and try out some of our new knowledge.

Putting Chapter 6 to Work

1. a. Explain why the IQ score is a continuous variable.
   b. What are the mean and the standard deviation for the distribution of IQ scores? SAT scores? Standard scores?
   c. Express, algebraically or as an equation, the relationship between standard scores and IQ scores and between standard scores and SAT scores.
   d. What standard score is 2 standard deviations above the mean? What IQ score is 2 standard deviations above the mean? What SAT score is 2 standard deviations above the mean?
   e. Compare the information about percentage of distribution shown in Figure 2.2 above with the empirical rule studied in Chapter 2. Explain the similarities.

2. Let’s take a second look at the normally distributed IQ scores illustrated in “Intelligence Scores.”
   a. How is an IQ score converted to a standard score?
   b. What is the standard score for an IQ score of 90? 110? 120?
   c. What is the standard score for an SAT score of 465? 575? 650?
   Using Figure 2.2 with the empirical rule:
   d. What percentage of IQ scores is greater than 132?
   e. What percentage of SAT scores is less than 700?
   Using Table 3 in Appendix B:
   f. What is the probability that an IQ score is greater than 132?
   g. What is the probability that an SAT score is less than 700?
   h. Compare your answers to parts f and g with your answers to parts d and e that used the empirical rule and Figure 2.2. Explain any similarities.
   i. What proportion of the IQ scores fall within the range of 80 to 120?
   j. What proportion of the IQ scores exceed 125?
   k. What percentage of the SAT scores are below 450?
   l. What percentage of the SAT scores are above 575?
   m. What SAT score is at the 95th percentile? Explain what this means.

Your Study

3 Intelligence Tests

The Wechsler Tests. Wechsler Adult Intelligence Scale-Revised, WAIS-R, and Wechsler Intelligence Scale for Children, WISC-III, are a widely used alternative to the Stanford-Binet. The Wechsler tests rate performance (non-verbal) intelligence in addition to verbal intelligence, and can be broken down to reveal strengths and weaknesses in various areas.

Based on scores from a large number of randomly selected people, IQ ranges have been classified as shown in Table 10-4. A look at the percentages reveals a definite pattern. The distribution of IQ approximates a normal curve, in which the majority of scores fall close to the average, with fewer at the extremes.
Chapter 6 Project

TABLE 10-4
Distribution of Adult IQ Scores on WAIS-R

<table>
<thead>
<tr>
<th>IQ</th>
<th>Description</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above 130</td>
<td>Very superior</td>
<td>2.2</td>
</tr>
<tr>
<td>120-129</td>
<td>Superior</td>
<td>6.7</td>
</tr>
<tr>
<td>110-119</td>
<td>Bright normal</td>
<td>16.1</td>
</tr>
<tr>
<td>90-109</td>
<td>Average</td>
<td>50.0</td>
</tr>
<tr>
<td>80-89</td>
<td>Dull normal</td>
<td>16.1</td>
</tr>
<tr>
<td>70-79</td>
<td>Borderline</td>
<td>6.7</td>
</tr>
<tr>
<td>Below 70</td>
<td>Mentally retarded</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Source: Dennis Coon, Essentials of Psychology, Exploration and Application, 8th ed. (Belmont, CA: Wadsworth, 1999)

a. Use the information in Table 10-4 and estimate the standard deviation for adult WAIS-R scores. Use at least two different pieces of information to obtain two separate estimates. Determine your answer.

b. Does the IQ score discussed here seem to have a normal distribution? Give reasons to support your answer.

c. What percentage of the adult population has “superior” intelligence?

d. What is the probability of randomly selecting one person from this population who is classified below “average”?

e. What IQ score is at the 95th percentile? Explain what this means.
The U.S. Census and Sampling It

According to the 2000 census, the U.S. population consists of more than 275 million people. We read and hear about this population often; the news media reports on results of samples nearly every day. One of the variables of interest to many is the “age” of Americans.

According to the 2000 census, the approximately 275 million Americans have a mean age of 36.5 years and a standard deviation of 22.5 years. The ages are distributed as shown in the accompanying histogram.

A census in the United States is done only every 10 years. It is an enormous and overwhelming job, but the information that is obtained is vital to our country’s organization and structure. Issues come up and times change; information is needed, and a census is impractical. This is where a representative sample comes in.

The fundamental goal of a survey is to come up with the same results that would have been obtained had every single member of a population been interviewed. For national Gallup polls, in other words, the objective is to present the opinions of a sample of people which are exactly the same opinions that would have been obtained had it been possible to interview all adult Americans in the country.

The key to reaching this goal is a fundamental principle called equal probability of selection, which states that if every member of a population has an equal probability of being selected in a sample, then that sample will be representative of the population.

Thus, it is Gallup’s goal in selecting samples to allow every adult American an equal chance of falling into the sample. How that is done, of course, is the key to the success or failure of the process.

Source: http://www.gallup.com/
As noted previously, the fundamental goal of a survey is to come up with the same results that we would have obtained had we interviewed every person of the population. Knowing that interviewing every person of a population is nearly impossible for most populations promotes the importance of a good representative sample. In addition, we now have the sampling distribution of sample means and the central limit theorem to help us make predictions about the population by using the sample. Putting Chapter 7 to Work will help us put these new concepts together.

Putting Chapter 7 to Work

Suppose a random sample of 100 ages was taken from 2000 census distribution. [CP07-1]

1. a. How would you graphically describe the 100 “ages” in the preceding random sample taken from the 2000 census distribution? Construct the graph.
   b. Using the graph that you constructed in part a, describe the shape of the distribution of sample data.
   c. How well did the sample describe the population of ages from the 2000 census shown in “U.S. Population—Census 2000” graphic? Explain using the graphical displays.
   d. How would you describe the preceding “ages” sample data numerically? Calculate the statistics.
   e. How well do the statistics calculated in part d compare with the parameters from the 2000 census given in “The U.S. Census and Sampling It”?

2. a. How would you numerically describe the 100 “ages” in the preceding random sample taken from the 2000 census distribution? Calculate the statistics.
   b. How well do the statistics calculated in part a compare with the parameters from the 2000 census? Be specific.
   c. If another sample was collected, would you expect the same results? Explain.

3. A second sample of 100 ages as been collected from the U.S. 2000 census and is listed here. [CP07-3]

   a. How would you describe the preceding “ages” sample data graphically? Construct the graph.
   b. Using the graph that you constructed in part a, describe the shape of the distribution of sample data.
   c. How well did the sample describe the population of ages from the 2000 census shown in “U.S. Population—Census 2000” graphic? Explain using the graphical displays.
   d. How would you describe the preceding “ages” sample data numerically? Calculate the statistics.
   e. How well do the statistics calculated in part d compare with the parameters from the 2000 census given in “The U.S. Census and Sampling It”?
   f. How does your graphical display and statistics compare with those constructed and calculated in questions 1 and 2 using a different sample of 100 ages?
   g. Is the distribution of ages for the population of Americans normal? Is it approximately normal?
   h. Will the SDSM apply to samples taken from this population? Explain.
   i. Will the CLT apply to samples taken from this population? Explain.
   j. Describe the SDSM for samples of size 100. Be sure to include center, spread, and shape.
   k. Compare your results in parts a and d with the theoretical answers in part j. Be sure to include center, spread, and shape.
   l. Describe the SDSM for samples of size 30. Be sure to include center, spread, and shape.
   m. Describe the sampling SDSM for samples of size 1000. Be sure to include center, spread, and shape.
   n. Relate your findings in parts j, l, and m to the SDSM and the CLT.
Your Study

4 Skillbuilder Applet Exercise simulates taking samples of size 50 from the population of American ages from the 2000 census, where $\mu = 36.5$ and $\sigma = 22.5$ and the shape is skewed right.

a. Click “1” for “# Samples.” Note the 50 data values and their mean. Change “slow” to “batch” and take at least 1000 samples of size 50.

b. What is the mean of the sample means? How close is it to the population mean?

c. What is the standard deviation of the sample means?

d. Based on the SDSM (as described in Section 7.3), what should you expect for the standard deviation of sample means? How close was your standard deviation from part c?

e. What shape is the histogram of the 1000 means?

f. Relate your findings to the SDSM and the CLT.
Part 2 Project

Working with Your Own Data

Putting Probability to Work

The sampling distribution of sample means and the central limit theorem are very important to the development of the rest of this course. The proof, which requires the use of calculus, is not included in this textbook. However, the truth of the SDSM and the CLT can be demonstrated both theoretically and by experimentation. The following activities will help to verify both statements.

A The Population
Consider the theoretical population that contains the three numbers 0, 3, and 6 in equal proportions.

1. a. Construct the theoretical probability distribution for the drawing of a single number, with replacement, from this population.
   b. Draw a histogram of this probability distribution.
   c. Calculate the mean, \( \mu \), and the standard deviation, \( \sigma \), for this population.

B The Sampling Distribution, Theoretically
Let’s study the theoretical sampling distribution formed by the means of all possible samples of size 3 that can be drawn from the given population.

2. Construct a list showing all the possible samples of size 3 that could be drawn from this population. (There are 27 possibilities.)
3. Find the mean for each of the 27 possible samples listed in answer to question 2.
4. Construct the probability distribution (the theoretical sampling distribution of sample means) for these 27 sample means.
5. Construct a histogram for this sampling distribution of sample means.
6. Calculate the mean \( \mu_x \) and the standard error of the mean \( \sigma_x \) using the probability distribution found in question 4.
7. Show that the results found in questions 1c, 5, and 6 support the three claims made by the sampling distribution of sample means and the central limit theorem. Cite specific values to support your conclusions.

C The Sampling Distribution, Empirically
Let’s now see whether the sampling distribution of sample means and the central limit theorem can be verified empirically; that is, does it hold when the sampling distribution is formed by the sample means that result from several random samples?

8. Draw a random sample of size 3 from the given population. List your sample of three numbers and calculate the mean for this sample.
   You may use a computer to generate your samples. You may take three identical “tags” numbered 0, 3, and 6, put them in a “hat,” and draw your sample using replacement between each drawing. Or you may use dice; let 0 be represented by 1 and 2; 3, by 3 and 4; and 6, by 5 and 6. You may also use random numbers to simulate the drawing of your samples. Or you may draw your sample from the list of random samples at the end of this section. Describe the method you decide to use. (Ask your instructor for guidance.)

9. Repeat question 8 forty-nine more times so that you have a total of 50 sample means that have resulted from samples of size 3.
10. Construct a frequency distribution of the 50 sample means found in questions 8 and 9.
11. Construct a histogram of the frequency distribution of observed sample means.
12. Calculate the mean \( \bar{x} \) and standard deviation \( s_x \) of the frequency distribution formed by the 50 sample means.
Part 2 Project

13. Compare the observed values of $\bar{x}$ and $s_x$ with the values of $\mu_x$ and $\sigma_x$. Do they agree? Does the empirical distribution of $\bar{x}$ look like the theoretical one?

Here are 100 random samples of size 3 that were generated by computer: [DS2]

<table>
<thead>
<tr>
<th>630</th>
<th>030</th>
<th>660</th>
<th>336</th>
<th>663</th>
<th>633</th>
</tr>
</thead>
<tbody>
<tr>
<td>003</td>
<td>306</td>
<td>300</td>
<td>636</td>
<td>030</td>
<td>663</td>
</tr>
<tr>
<td>666</td>
<td>030</td>
<td>633</td>
<td>330</td>
<td>330</td>
<td>333</td>
</tr>
<tr>
<td>333</td>
<td>306</td>
<td>660</td>
<td>336</td>
<td>006</td>
<td>063</td>
</tr>
<tr>
<td>666</td>
<td>006</td>
<td>330</td>
<td>066</td>
<td>003</td>
<td>663</td>
</tr>
<tr>
<td>006</td>
<td>006</td>
<td>666</td>
<td>636</td>
<td>660</td>
<td>300</td>
</tr>
<tr>
<td>366</td>
<td>630</td>
<td>363</td>
<td>300</td>
<td>336</td>
<td>060</td>
</tr>
<tr>
<td>300</td>
<td>036</td>
<td>633</td>
<td>606</td>
<td>336</td>
<td>603</td>
</tr>
<tr>
<td>036</td>
<td>363</td>
<td>663</td>
<td>660</td>
<td>333</td>
<td>300</td>
</tr>
<tr>
<td>630</td>
<td>660</td>
<td>030</td>
<td>660</td>
<td>366</td>
<td>036</td>
</tr>
<tr>
<td>633</td>
<td>030</td>
<td>660</td>
<td>663</td>
<td>660</td>
<td>303</td>
</tr>
<tr>
<td>363</td>
<td>360</td>
<td>006</td>
<td>033</td>
<td>366</td>
<td>036</td>
</tr>
<tr>
<td>060</td>
<td>600</td>
<td>060</td>
<td>066</td>
<td>033</td>
<td>036</td>
</tr>
<tr>
<td>336</td>
<td>333</td>
<td>336</td>
<td>636</td>
<td>333</td>
<td>366</td>
</tr>
<tr>
<td>633</td>
<td>300</td>
<td>306</td>
<td>603</td>
<td>366</td>
<td>603</td>
</tr>
<tr>
<td>033</td>
<td>630</td>
<td>036</td>
<td>036</td>
<td>033</td>
<td>036</td>
</tr>
</tbody>
</table>
Chapter 8 Project

Were They Shorter Back Then?

WERE THEY SHORTER BACK THEN?

The average height for an early 17th-century English man was approximately 5' 6". For 17th-century English women, it was about 5' 1/2". While average heights in England remained virtually unchanged in the 17th and 18th centuries, American colonists grew taller. Averages for modern Americans are just over 5' 9" for men, and about 5' 3 3/4" for women. The main reasons for this difference are improved nutrition, notably increased consumption of meat and milk, and antibiotics.

Source: http://www.plimoth.org/library/short.htm

The National Center for Health Statistics (NCHS) provides statistical information that guides actions and policies to improve the health of the American people. Recent data from NCHS give the average height of females in the United States to be 63.7 inches, with a standard deviation of 2.75 inches.

A random sample of 50 females from the health profession yielded the following height data.

<table>
<thead>
<tr>
<th>Heights (inches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65.0 66.0 64.0 67.0 59.0 69.0 66.0 69.0 64.0 61.5</td>
</tr>
<tr>
<td>63.0 62.0 63.0 64.0 72.0 66.0 65.0 64.0 67.0 68.0</td>
</tr>
<tr>
<td>70.0 63.0 63.0 68.0 58.0 60.0 63.5 66.0 64.0 62.0</td>
</tr>
<tr>
<td>64.5 69.0 63.5 69.0 62.0 58.0 66.0 68.0 59.0 56.0</td>
</tr>
<tr>
<td>64.0 66.0 65.0 69.0 67.0 66.5 67.5 62.0 70.0 62.0</td>
</tr>
</tbody>
</table>

Putting Chapter 8 to Work

1. Do you expect the mean of this random sample of 50 females to be exactly equal to the population mean of 63.7 inches given by NCHS? If the sample mean is greater than 63.7 inches, does it mean that we are even taller today?

2. [CP08] a. What population was sampled to obtain the height data listed?
   b. Describe the sample data using the mean and standard deviation, plus any other statistics that help describe the sample. Construct a histogram and comment on the shape of the distribution.

3. a. How is the distribution of the sample height related to the distribution of the population and the sampling distribution of sample means?
   b. Using the techniques in Chapter 7, find the limits that would bound the middle 90% of the sampling distribution of sample means for samples of size 50 randomly selected from the population of female heights with a known mean of 63.7 inches and a standard deviation of 2.75 inches.
   c. On the histogram drawn in question 2b, draw a vertical line at the population mean of 63.7 and draw a horizontal line segment showing the interval found in part b. Does the sample mean found in question 2b fall in the interval? Explain what this means.
   d. Using the techniques of Chapter 7, find $P(\bar{x} \geq 64.7)$ for a random sample of 50 drawn from a population with a known mean of 63.7 inches and a standard deviation of 2.75 inches. Explain what the resulting value means.
   e. Does the sample of 50 height data values appear to belong to the population described by the NCHS? Explain.
Chapter 8 Project

4 a. Are the assumptions of the confidence interval and hypothesis test methods of this chapter satisfied? Explain.

b. Using the sample data and a 95% level of confidence, estimate the mean height of females in the health profession. Use the given population standard deviation of 2.75 inches.

c. Test the claim that the mean height of females in the health profession is different from 63.7 inches, the mean height for all females in the United States. Use a 0.05 level of significance.

d. On the same histogram used in part b of question 2:
   (i) Draw a vertical line at the hypothesized population mean value, 63.7.
   (ii) Draw a horizontal line segment showing the 95% confidence interval found in part b.

e. Does the mean $\mu = 63.7$ fall in the interval? Explain what this means.

f. Describe the relationship between the two lines drawn on your graph for part c of question 3 and the two lines drawn for part d of this exercise.

g. On the basis of the results obtained earlier, does it appear that the females in this study, on average, are the same height as all females in the United States as reported by the NCHS? Explain.

Your Study

5 Design your own study of female heights. Define a specific population that you will sample, describe your sampling plan, collect your data, and answer part b of question 2 and parts a, b, c, and g of question 4 replacing health profession with your particular population. Discuss the differences and similarities between your sample and the population and between your sample and the sample of 50 female health professionals.
Chapter 9 Project

Get Enough Daily Exercise?

GOOD NEWS FOR WOMEN

The National Women’s Health Information Center reports, “It’s never too late to start an active lifestyle.” No matter how old you are, how unfit you feel, or how long you have been inactive, research shows that starting a more active lifestyle now through regular, moderate-intensity activity can make you healthier and improve your quality of life. Here’s what you should do:

<table>
<thead>
<tr>
<th>If:</th>
<th>Then:</th>
</tr>
</thead>
<tbody>
<tr>
<td>You do not currently engage in regular physical activity.</td>
<td>you should begin by incorporating a few minutes of physical activity into each day, gradually building up to 30 minutes or more of moderate-intensity activities.</td>
</tr>
<tr>
<td>You are now active, but at less than the recommended levels,</td>
<td>you should strive to adopt more consistent activity:</td>
</tr>
<tr>
<td>You currently engage in moderate-intensity activities for at least 30 minutes on 5 days or more of the week,</td>
<td></td>
</tr>
<tr>
<td>You currently regularly engage in vigorous-intensity activities 20 minutes or more on 3 days or more of the week,</td>
<td>you may achieve even greater health benefits by increasing the time spent or intensity of those activities.</td>
</tr>
<tr>
<td>Scientific evidence to date supports the statements in this table.</td>
<td>you should continue to do so.</td>
</tr>
</tbody>
</table>

Source: Centers for Disease Control and Prevention, http://www.4woman.gov/

The article recommends different amounts and different levels of activity depending on a woman’s current level of activity. Some of the recommendations require as little as 60 minutes of exercise a week. The data values that follow are from a study surveying cardiovascular technicians (individuals who perform various cardiovascular diagnostic procedures) as to their own physical exercise per week, measured in minutes. [CP09]

<table>
<thead>
<tr>
<th>Activity Duration</th>
<th>Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>3</td>
</tr>
<tr>
<td>60</td>
<td>2</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>80</td>
<td>1</td>
</tr>
</tbody>
</table>

Do the technicians, on average, appear to exercise at least 60 minutes per week?
Chapter 9 Project

Putting Chapter 9 to Work

1. Consider the preceding sample data.
   a. What is the population parameter of interest?
   b. Construct a histogram of the data.
   c. What name would you give to the shape of the histogram in part b?
   d. Would you say the histogram in part b suggests the variable, amount of time, does not have a normal distribution?

2. Based on the data for the cardiovascular technicians:
   a. Find the mean and standard deviation for the amount of time the cardiovascular technicians exercised per week.
   b. How would you estimate the mean amount of time exercised per week by all cardiovascular technicians?
   c. Does it appear that cardiovascular technicians exercise “at least 60 minutes per week”? Justify your answer.

3. a. What evidence do you have to show that the assumption of normality is reasonable? Explain.
   b. Estimate the mean amount of weekly exercise time for all cardiovascular technicians using a point estimate and a 95% confidence interval.
   c. The “Good News for Women” article in “Get Enough Daily Exercise?” says that people should exercise at least 60 minutes a week. Based on the data from the study, determine whether the technicians exercise at least 60 minutes a week. Use a 0.05 level of significance.

Your Study

4. a. Define the population whose amount of exercise time per week you would be interested in investigating.
   b. Collect the “times” from a sample of 40 members of your population.
   c. Find the mean and standard deviation for the amount of time exercised per week by the members of your sample.
   d. Construct a graph displaying the distribution of your data.
   e. Estimate the mean amount of weekly exercise time for your population using a point estimate and a 95% confidence interval.
   f. The “Good News for Women” article says that people (can?) should exercise at least 60 minutes a week. Does it appear that the members of your sample exercise “at least 60 minutes per week”? Use a 0.05 level of significance. Justify your answer.
   g. Did your data satisfy the assumptions? Explain.
Chapter 10 Project

Students, Credit Cards, and Debt

We all know, “College is expensive” and “Credit cards are readily available.” We also know and believe that young adults need experience at handling their own finances. But two old adages, “Buyers beware” and “Know the facts” probably should each play a role in a college student’s approach to credit card use. Below is a portion of the report Nellie Mae published in 2002.

CREDIT CARDS AND DEBT

Not surprisingly, the freshman population has a lower overall percentage of credit cards and lower debt levels on their cards than students in upper classes. However, more than half of all freshmen (54%) had at least one credit card, with the average number of cards being 2.5; among those who have credit cards, 26% have four or more. Freshman debt levels are also lower than the overall counts in all categories. Their median debt amount is $901, lower than the overall median of $1770; their average balance is $1533 versus $2327 overall; those with balances exceeding $7000 account for only 4% of freshmen as opposed to 6% overall; and those with high-level balances between $3000 and $7000 account for 8% of freshmen compared with 21% overall.

As students progress through their 4 years (or more) in college, there is a steady increase in credit card usage rates and balances each year. By graduation, most students have more than doubled their average debt and almost tripled the number of cards they hold. Most dramatic, however, is the 70% jump occurring between freshman and sophomore year in the percentage of students with at least one card—from 54% to 92% of the total population.

Once freshmen arrive on campus, there are many tempting incentives to sign up for new credit cards and many opportunities to use them. The fact that the average number of cards per student continues to increase is not surprising. The proliferation of on-campus, mail, and Internet offers of free gifts, bonus airline miles, and low introductory rates for each new card is difficult for students to resist.

The data that follow are from random samples of 200 freshman and 200 sophomore college students who were asked, “Do you have your own credit card?” A total of 97 freshman and 187 sophomores answered that they had one or more credit cards with their name on it. The first 40 freshmen and the first 44 sophomores answering “yes” were then asked for their “current total credit card debt balance.” The total credit card debt balances are listed here.

CREDIT CARD USAGE BY GRADE LEVEL

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage Who Have Credit Cards</th>
<th>Average Number of Credit Cards</th>
<th>Percentage Who Have 4 or More Cards</th>
<th>Average Credit Card Debt</th>
<th>Median Credit Card Debt</th>
<th>Percentage with Balances between $3000 and $7000</th>
<th>Percentage with Balances Exceeding $7000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh</td>
<td>54%</td>
<td>2.5</td>
<td>26%</td>
<td>$1533</td>
<td>$901</td>
<td>8%</td>
<td>4%</td>
</tr>
<tr>
<td>Soph</td>
<td>92%</td>
<td>3.67</td>
<td>44%</td>
<td>$1825</td>
<td>$1564</td>
<td>18%</td>
<td>4%</td>
</tr>
<tr>
<td>Junior</td>
<td>87%</td>
<td>4.5</td>
<td>50%</td>
<td>$2705</td>
<td>$1872</td>
<td>24%</td>
<td>7%</td>
</tr>
<tr>
<td>Senior</td>
<td>96%</td>
<td>6.13</td>
<td>66%</td>
<td>$3262</td>
<td>$2185</td>
<td>31%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Source: Undergraduate Students and Credit Cards, published April 2002, Nellie Mae, 50 Brantree Hill Park, Suite 300, Braintree, MA 02184, 781-849-1325; http://www.nelliemae.com
Chapter 10 Project

As a way of assessing the statistical techniques for two populations that we have learned in this chapter, let’s return to “Students, Credit Cards, and Debt.” Credit card companies are notorious for enticing college students to sign up for new credit cards. The convenience of the credit cards along with money-handling inexperience lead to huge debts for many college students. Is there a significant difference between freshmen and sophomore college students with respect to credit cards and debt? Let’s investigate.

Putting Chapter 10 to Work

1. Consider the sample data above in “Students, Credit Cards, and Debt.”

   a. What is the population of interest?
   
   b. What percentage of each group has their own credit card? How does this compare with the findings reported by the Nellie Mae organization?
   
   c. Describe the shape of the distribution you believe total credit card debt will display. Explain.
   
   d. Construct a histogram of total credit card debt for each class. Use the same class intervals for both histograms. Compare your findings to your thoughts in part c.

2. How do credit card debts for freshmen and sophomores compare? Using the two sets of sample data above:

   a. Find the proportion of each sample that have at least one credit card.
   
   b. Find the point estimate for the difference between the two proportions.
   
   c. Are the assumptions for making inferences about the difference between two proportions satisfied? Explain.
   
   d. Find the 95% confidence interval for the difference between proportion of sophomores and freshmen having their own credit cards.
   
   e. Find the mean credit card debt for the freshmen and sophomores in the samples.
   
   f. Draw dotplots for the amount of debt for both groups using a common scale. Interpret what the dotplots are showing you, including shape, center, and spread.
   
   g. Find the point estimate for difference between two means.
   
   h. Check the assumptions for normality for both sets of credit card debt. Verify.
   
   i. What effect does your answer in part h have on answering the question: Is the difference between the mean credit card debt for sophomores and freshmen in part g significantly greater than the $292 difference reported by Nellie Mae above?
   
   j. Based on your finding in parts a–i, compare and contrast the spending habits of freshman and sophomore college students.

Your Study

3. Design your own study involving two populations.

   a. Determine a set of questions that compare the means, proportions, or variances of two populations of interest to you. You might consider two different class levels as the two populations and questions
similar to the following: Is there a difference between their mean cost of books and supplies for a semester? Is there a difference between the proportion of those with credit cards of their own? Is there a difference between the proportion of those with four or more credit cards of their own?

b. Define two specific populations that you will sample, describe your sampling plan, and collect the data needed to answer your questions.

c. Discuss any differences and similarities between your study and “Students, Credit Cards, and Debt.”
Part 3 Project

Working with Your Own Data

History contains many stories about consumers and the various products they purchase. An exhibit at the Boston Museum of Science tells such a mathematician–baker story. A man named Poincaré bought one loaf of bread daily from his local baker, a loaf that was supposed to weigh 1 kilogram. After a year of weighing and recording the weight of each loaf, Poincaré found a normal distribution with a mean of 950 grams. The police were called and the baker was told to behave himself; however, a year later Poincaré reported that the baker had not reformed and the police confronted the baker again. The baker questioned, “How could Poincaré have known that we always gave him the largest loaf?” Poincaré then showed the police the second year of his record, a bell-shaped curve with a mean of 950 grams but truncated on the left side.

As consumers, we all purchase many bottled, boxed, canned, and packaged products. Seldom, if ever, do any of us question whether or not the content is really the amount stated on the container. Here are some content listings found on containers we purchase:

- 28 FL OZ (1 PT 12 OZ) 750 ml
- 5 FL OZ (148 ml) 32 FL OZ (1 QT) 0.95 l
- NET WT 10 OZ NET WT 3 3/4 OZ
- 283 GRAMS 106g—48 tea bags
- 140 1-PLY NAPKINS 77 SQ FT—92 TWO-PLY SHEETS—11 × 11 IN.

Have you ever wondered, “Am I getting the amount that I am paying for?” And if this thought did cross your mind, did you attempt to check the validity of the content claim? The following article appeared in the Times Union of Rochester, New York, in 1972.

MILK FIRM ACCUSED OF SHORT MEASURE

The processing manager of Dairylea Cooperative, Inc., has been named in a warrant charging that the cooperative is distributing cartons of milk in the Rochester area containing less than the quantity represented.

... an investigator found shortages in four quarts of Dairylea milk purchased Friday. Asst. Dist. Atty. Howard R. Relin, who issued the warrant, said the shortages ranged from 1 1/2 to 1 1/2 ounces per quart. A quart of milk contains 32 fluid ounces.

... the state Agriculture and Markets Law ... provides that a seller of a commodity shall not sell or deliver less of the commodity than the quantity represented to be sold.

... the purpose of the law under which ... the dairy is charged is to ensure honest, accurate, and fair dealing with the public. There is no requirement that intent to violate the law be proved, he said.

Source: Times Union, Rochester, NY, February 16, 1972

This situation poses a very interesting legal problem: there is no need to show intent to “short the customer.” If caught, violators are fined automatically and the fines are often quite severe.

A A High-Speed Filling Operation

A high-speed piston-type machine used to fill cans with hot tomato juice was sold to a canning company. The guarantee stated that the machine would fill 48-oz cans with a mean amount of 49.5 oz, a standard deviation of 0.072 oz, and a maximum spread of 0.282 oz while operating at a rate of filling 150 to 170 cans per minute. On August 12, 1994, a sample of 42 cans was gathered and the following weights were recorded. The weights, measured to the nearest 1/2 oz, are recorded as variations from 49.5 oz.
Part 3 Project

[DS3]

1. Calculate the mean $\overline{x}$, the standard deviation $s$, and the range of the sample data.
2. Construct a histogram picturing the sample data.
3. Does the amount of fill differ from the prescribed 49.5 oz at the $\alpha = 0.05$ level? Test the hypothesis that $\mu = 49.5$ against an appropriate alternative.
4. Does the amount of variation, as measured by the range, satisfy the guarantee?
5. Assuming that the filling machine continues to fill cans with an amount of tomato juice that is distributed normally and the mean and standard deviation are equal to the values found in question 1, what is the probability that a randomly selected can will contain less than the 48 oz claimed on the label?
6. If the amount of fill per can is normally distributed and the standard deviation can be maintained, find the setting for the mean value that would allow only 1 can in every 10,000 to contain less than 48 oz.

B Your Own Investigation
Select a packaged product that has a quantity of fill per package that you can and would like to investigate.
1. Describe your selected product, including the quantity per package, and describe how you plan to obtain your data.
2. Collect your sample of data. (Consult your instructor for advice on size of sample.)
3. Calculate the mean $\overline{x}$ and the standard deviation $s$ for your sample data.
4. Construct a histogram or stem-and-leaf diagram picturing the sample data.
5. Does the mean amount of fill agree with the amount given on the label? Test using $\alpha = 0.05$.
6. Assume that the item you selected is filled continually. The amount of fill is normally distributed, and the mean and standard deviation are equal to the values found in question 3. What is the probability that one randomly selected package contains less than the prescribed amount?
Chapter 11 Project

Cooling a Great Hot Taste

If you like hot foods, you probably have a preferred way to “cool” your mouth after eating a delicious spicy favorite. Some of the more common methods used by people are drinking water, milk, soda, or beer or eating bread or other food. There are even a few people who prefer not to cool their mouth on such occasions and therefore do nothing. The graphic shown here lists the top six ways adults say they cool their mouths after eating hot sauce.

Putting Out The Fire

Top six ways American adults say they cool their mouths after eating hot sauce:

<table>
<thead>
<tr>
<th>Method</th>
<th>Water</th>
<th>Milk</th>
<th>Soda</th>
<th>Beer</th>
<th>Bread</th>
<th>Other</th>
<th>Nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>73</td>
<td>35</td>
<td>20</td>
<td>19</td>
<td>29</td>
<td>11</td>
<td>13</td>
</tr>
</tbody>
</table>

Source: Data from Anne R. Carey and Suzy Parker, © 1995 USA Today.

Two hundred adults professing to love hot spicy food were asked to name their favorite way to cool their mouth after eating food with hot sauce. Following is the summary of the resulting sample. [CP11]

Putting Chapter 11 to Work

1 Referring to the sample of 200 adults collected in “Cooling a Great Hot Taste”: 
   a. What information was collected from each adult in the sample?
   b. Define the population and the variable involved in the sample.
   c. Using the sample data, calculate percentages for the various methods of cooling one’s mouth.
   d. How do the sample percentages compare with the graphic percentages?

2 Referring to the sample of 200 adults collected in “Cooling a Great Hot Taste”, how similar is the distribution in the sample to the distribution of percentages in the graphic?
   a. Construct a horizontal bar graph of the 200 adults using relative frequency for the horizontal scale.
   b. Superimpose the bar graph from “Putting Out The Fire” on the bar graph in part a.
   c. Would you say the sample’s distribution looks “similar to” or “quite different than” the distribution shown in the “Putting Out The Fire” graph? Explain your answer.
Chapter 11 Project

3  a. Does the sample show a distribution that is significantly different from the distribution shown in the “Putting Out The Fire” graph? Use $\alpha = 0.05$.
   
b. Write a paragraph (50 + words) describing why the statistical method used in part a is appropriate for this set of data.
   
c. Write a paragraph (50 + words) describing the meaning of the assumptions and the results of the statistical procedure chosen in part b.

Your Study

4 Design your own study for “favorite way people cool their mouth after eating something hot.”
Chapter 12 Project

Time Spent Commuting to Work

How much time did America’s workforce spend getting to their jobs this morning? How much time did your parents spend commuting to work this morning? Does everybody spend the same amount of time? What was the mean amount of time spent commuting to work this morning by people in Boston? What was the mean amount of time spent commuting to work this morning by people in Dallas? Do you think that the city will have any effect on the amount of time spent commuting to work this morning? The graphic “Longest Commute To Work” seems to suggest that some cities have longer commuting times than others. From studying previous chapters, we know that the statistics from different samples, even if drawn from the same population, vary. The question that might be asked here is, “Is the variation between the samples greater than would be expected if the samples were all drawn from one population?"

Longest Commute To Work

Among U.S. cities with populations of 250,000 or more, New York City has the longest commute. The average one-way commuting time for all large U.S. cities is 24.3 minutes.

Riverside 31.2 min.
Newark 31.5 min.
Chicago 33.2 min.
New York 38.3 min.

To compare the commuting time in various locations, independent and random samples were obtained in each of six different U.S. cities from workers who commute to work during the 8:00 AM rush hour.

| TABLE 1 One-Way Travel to Work (in minutes) [CP12] |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Atlanta | Boston | Dallas | Philadelphia | Seattle | St. Louis | Atlanta | Boston | Dallas | Philadelphia | Seattle | St. Louis |
| 29 | 18 | 42 | 29 | 30 | 15 | 37 | 32 | 20 | 42 | 30 | 33 |
| 21 | 37 | 25 | 20 | 19 | 54 | 26 | 34 | 26 | 48 | 35 |
| 20 | 37 | 36 | 33 | 31 | 42 | 15 | 25 | 32 | 39 | 23 |

Do you think that what city you live in has an effect on the average commute time? The graphic “Longest Commute To Work” seems to suggest that the city has an effect on the commute time. Let’s investigate this question using the sample data from the six different U.S. cities.

Putting Chapter 12 to Work

1. a. Construct a graphic representation of the Table 1 data using six side-by-side dotplots.

b. Visually estimate the mean commute time for each city and locate it with an X.

c. Does it appear that the city has an effect on the average amount of time spent by workers who commute to work during the 8:00 AM rush hour? Explain.

d. Does it visually appear that the city has an effect on the variation in the amount of time spent by workers who commute to work during the 8:00 AM rush hour? Explain.
2 a. Calculate the mean commute time for each city in Table 12.1.

b. Does there seem to be a difference among the mean one-way commute times for these six cities?

c. Calculate the standard deviation for each of the six cities’ commute times.

d. Does there seem to be a difference among the standard deviations of the one-way commute times for these six cities?

3 Let \( x \) = one-way travel to work in minutes. Independent and random samples were obtained in each of the six different U.S. cities from workers who commute to work during the 8:00 AM rush hour.

<table>
<thead>
<tr>
<th>City</th>
<th>Atlanta</th>
<th>Boston</th>
<th>Dallas</th>
<th>Philadelphia</th>
<th>Seattle</th>
<th>St. Louis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29</td>
<td>18</td>
<td>42</td>
<td>29</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>37</td>
<td>25</td>
<td>20</td>
<td>19</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>37</td>
<td>36</td>
<td>33</td>
<td>31</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>25</td>
<td>32</td>
<td>37</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>32</td>
<td>20</td>
<td>42</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>34</td>
<td>26</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Construct a side-by-side boxplot showing the six cities.

b. Does your graph show visual evidence suggesting that the city has an effect on the average morning commute time? Justify your answer.

c. Using the ANOVA technique learned in this chapter, do these data show sufficient evidence to claim that a person’s age has an effect on the average amount of time spent commuting to work? Use \( \alpha = 0.05 \).

d. Discuss the similarities and differences between results of your study and the results found in question 3.

Your Study

4 Design your own “commute time” study.

a. Define a specific population that you will sample based on four age groupings, describe your sampling plan, and collect a random sample of at least 25 observations.

b. Make a descriptive presentation of your sample data, using a chart or graph and a descriptive paragraph.

c. Using the ANOVA technique learned in this chapter, do these data show sufficient evidence to claim that a person’s age has an effect on the average amount of time spent commuting to work in the morning? Use \( \alpha = 0.05 \).

d. Discuss the similarities and differences between results of your study and the results found in question 3.
Wheat! Beautiful Golden Wheat!

Wheat is a common name for cereal grass of a genus of the grass family. It has been cultivated for food since prehistoric times and is one of our most important grain crops. The common types of wheat grown in the United States are spring wheat, planted in the spring for fall harvest, and winter wheat, planted in the fall for spring harvest. The main use of wheat is in flour for bread and other food products. It’s also used to a limited extent in the making of beer, whiskey, industrial alcohol, and other products.

THE U.S. WHEAT CROP FOR 2002 IS EXPECTED TO BE THE SMALLEST IN A QUARTER CENTURY

July 12, 2002 (EIRNS)—The area harvested in the United States this year for winter wheat (the predominant wheat variety in U.S. latitudes) is estimated to be only 29.8 million acres (12.06 million hectares)—the same as in 1917! (The United States harvested winter wheat area in recent years has been between 35 and over 40 million acres.) Farmers have abandoned large amounts of sown land because of drought and related pests and disease. Estimates now put the total U.S. wheat harvest (all types) this year at around 1.79 billion bushels (48.9 million metric tons), about the same as in 1974, and way down from the 64 million ton levels of recent yearly harvests. Western Canada could potentially harvest 19.7 million metric tons of wheat, down from the five-year average of 23.3 million tons, which itself has been declining.

In terms of world trade in basic foodstuffs, the United States and Canada are a major source of world wheat supplies—now severely contracted. Australia’s wheat output next season is expected to drop. Argentina is in turmoil. Only Europe (principally France) expects a good harvest. World wheat stocks are way down.

Source: http://committeeRepublicCanada.ca/

The business decisions a grain farmer makes are not as simple as the statistical relationship between the four variables listed in the table that follows. However, an understanding of the relationship between these variables is an important component of what a grain farmer needs to know in order to make decisions about how many acres to plant, what kind of grain to plant, and so on.

Twenty randomly selected wheat-producing counties in Kansas were identified and data were collected for these variables:

- **Planted** = 1000s of acres planted with winter wheat
- **Harvested** = 1000s of acres harvested (not all planted acres are harvested for a variety of reasons)
- **Yield** = bushels of wheat harvested per acre
- **Production** = 1000s of bushels of wheat harvested
The four variables defined in “Wheat! Beautiful Golden Wheat!” all seem to be interrelated. Upon completion of questions 1 and 2 below, one finds that a strong linear relationship exists between the amount planted and the amount harvested, as well as the amount harvested and the production. As one increased, so did the other. Intuitively, this makes sense. Let’s further our investigation of relationships with the combinations of planted with production and planted with yield to see if they also exist and have meaning.

**Putting Chapter 13 to Work**

1. Refer to the wheat crop data above.
   a. Construct a scatter diagram for Planted \((x)\) and Harvested \((y)\).
   b. Construct a scatter diagram of Harvested \((x)\) and Production \((y)\).
   c. Construct a histogram of Yield.
   d. Explain what is learned from parts a–c.

2. Refer to the wheat crop data above.
   a. Find the linear correlation coefficient and the equation for the line of best fit for Planted \((x)\) and Harvested \((y)\). What does the correlation coefficient tell you with respect to the variables?
   b. Find the linear correlation coefficient and the equation for the line of best fit for Harvested \((x)\) and Production \((y)\). What does the correlation coefficient tell you with respect to the variables?

3. Refer to the wheat crop data above.
   a. Construct a scatter diagram and find the linear correlation coefficient and the equation for the line of best fit for Planted \((x)\) and Production \((y)\).
   b. Find the 95% confidence interval for the true linear correlation coefficient between Planted and Production.
   c. Calculate the 95% confidence interval for the slope, \(\beta_1\), for the line of best fit relating Planted and Production.
   d. Predict the total production for a county with 125 thousand acres planted using a 95% prediction interval.
   e. Construct a scatter diagram and find the linear correlation coefficient and the equation for the line of best fit relating Planted \((x)\) and Yield \((y)\).
   f. Test the significance of the correlation between Planted and Yield. Use \(\alpha = 0.05\).
   g. Test the significance of the slope, \(\beta_1\), for the line of best fit relating Planted and Yield. Use \(\alpha = 0.05\).
   h. Estimate the mean yield for all counties with 100 thousand acres planted with a 95% confidence interval.

4. Design your own study to investigate another grain. Answer questions similar to those asked in question 3. Use the Internet and the U.S. Department of Agriculture’s National Agricultural Statistics Service website (http://www.usda.gov/) to find your data. The data sets posted will contain extra information that makes the data in its original format unusable. However, this unwanted information can easily be removed by deleting the unwanted rows and columns. (Note: http://www.aragriculture.org/ also has many data sets posted.)
Chapter 14 Project

Teenagers’ Attitudes

A national survey revealed that teenagers’ attitudes toward moral and social values are much more conventional than widely believed.

How Teenagers See Things

One could think of them as Generation “V”—for values. According to The Mood of American Youth study, today’s teens are neither as rebellious as adolescents in the 1970s nor as materialistic as those of the 1980s. What they want is not to change the world or to own a chunk of it, but to be happy. Among the teens’ greatest concerns: the decline in moral and social values.

The study, conducted by NFO Research, Inc., included 938 young people aged 13 to 17 who are representative of America’s adolescent population as a whole. Of those polled, 9 in 10 say they don’t drink or smoke, and 7 in 10 say religion is important in their lives. Most of the respondents respect their parents, get along well with them, and consider their rules strict but fair.

Teens’ Views on Contemporary Issues

Families and Children

<table>
<thead>
<tr>
<th>Choices</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teenagers are not prepared to have babies.</td>
<td>91%</td>
</tr>
<tr>
<td>A single parent can raise a family.</td>
<td>75%</td>
</tr>
<tr>
<td>I am very likely to raise my children differently than I was raised.</td>
<td>55%</td>
</tr>
</tbody>
</table>

In the Schools

<table>
<thead>
<tr>
<th>Choices</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local school officials should be able to censor the books and materials used in their schools.</td>
<td>47%</td>
</tr>
<tr>
<td>School prayer should not be permitted.</td>
<td>32%</td>
</tr>
</tbody>
</table>

Social Concerns

<table>
<thead>
<tr>
<th>Choices</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The “V” chip will unfairly censor what teens can watch on TV.</td>
<td>67%</td>
</tr>
<tr>
<td>It is important to control information on the Internet.</td>
<td>60%</td>
</tr>
</tbody>
</table>

. . . And What about the Government?

<table>
<thead>
<tr>
<th>Choices</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government spending on AIDS research should be increased.</td>
<td>83%</td>
</tr>
<tr>
<td>Adequate health care for all should be provided through a national health plan.</td>
<td>81%</td>
</tr>
<tr>
<td>Being rich is necessary to get elected to high office.</td>
<td>55%</td>
</tr>
</tbody>
</table>

Part of the survey asked the teens to indicate the one thing they wanted most from life. The following table lists the choices and the percentage of teens selecting each.

The One Thing Teens Want Most from Life* [CP14]

<table>
<thead>
<tr>
<th>Choices</th>
<th>All (%)</th>
<th>Boys (%)</th>
<th>Girls (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>28</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td>Long, enjoyable life</td>
<td>16</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>Marriage and family</td>
<td>9</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Financial success</td>
<td>8</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Career success</td>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Religious satisfaction</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Love</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Personal success</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Personal contribution to society</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Friends</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Health</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

*Some teens didn’t respond, so figures don’t total 100%.

Source: © 1996 Dianne Hales. All rights reserved. From Parade Magazine, August 18, 1996, and reprinted with permission.
Putting Chapter 14 to Work

1. Consider the table “The One Thing Teens Want Most from Life” on page 37:
   a. Does it appear that teenage boys and girls agree on what they want most from life?
   b. Construct a scatter diagram of Boys’ Percent Choices versus Girls’ Percent Choices.
   c. Does the graph constructed in part b support your answer in part a?

2. a. Construct a side-by-side bar chart of the Boys’ Percent Choices and the corresponding Girls’ Percent Choices.
   b. Based on the graph constructed in part a, does it appear that teenage boys and girls agree on what they want most from life?
   c. Explain why the chi-square test of homogeneity studied in Chapter 11 cannot be completed using this information.

3. a. Do responses from the boys and the girls have the same distribution? Use the Mann–Whitney U test and \( \alpha = 0.05 \).
   b. Rank the choices for the boys and girls separately of each other.
   c. Do the boys’ preferences correlate to the girls’ preferences? Use Spearman’s rank correlation to test at the 0.05 level of significance.
   d. Compare the results obtained in parts b and c.

Your Study

4. Define a population of your choice and randomly sample the following variables:
   Variable 1: Gender or class level (freshmen, sophomore, upper classes, etc.)
   Variable 2: Name “The One Thing Wanted Most from Life”; choose from happiness, long enjoyable life, marriage and family, financial success, career success, religious satisfaction, love, personal success, personal contribution to society, friends, health, or education
   a. Rank the choices for each level of variable 1.
   b. Do the responses from the levels of variable 1 have the same distribution? Use the Mann–Whitney U test and \( \alpha = 0.05 \)
   c. Do the levels of the responses correlate to each other? Use Spearman’s rank correlation to test at the 0.05 level of significance.
   d. Write a paragraph comparing and contrasting the results obtained in parts b and c.
Part 4 Project

Working with Your Own Data

The existence of bivariate data is commonplace in everyday life and there are multiple options for analyzing the relationship between the two variables. In Chapter 10, the relationship between the paired data was analyzed as paired differences. In Chapter 12, the analysis of variance methods tested for an effect that one variable might have on a second variable. In Chapter 13, the methods of correlation and regression were used to investigate the relationship between the variables in order to determine whether they have a mathematical relationship that can be approximated by means of a straight line. The following illustrates such a situation.

A The Age and Value of Peggy’s Car

Peggy would like to sell her 1994 Corvette, and she wants to determine an asking price for it in order to advertise. Her Corvette features the typical Corvette equipment with no customizing and is in average condition for a well-cared-for 1994 Corvette. She wants to advertise using an average asking price and expects to get an average price for it (average for a Corvette!) when it sells. Presently she must answer the question, “What is an average asking price for a 1994 Corvette?”

Inspection of many classified sections of newspapers turned up only three advertisements for 1994 Corvettes. The prices listed varied a great deal, and Peggy needed more information to determine her asking price. She decided to use the Internet and search for prices. She restricted her search to used Chevrolet Corvettes, 1988 to present, that were in good repair, not customized or show cars, and were being sold by their owner, not a dealer or an auction. Peggy collected the following data on January 8, 2003.

Peggy knows that the price she should ask, and the amount she receives, for her Corvette is affected by its age. The general questions are: “Is the effect of age on price predictable?” and “Can a meaningful relationship between the age and the typical asking price for used Corvettes be established?”

Independent variable, \(x\): The age of the car as measured in years and defined by

\[ x = (\text{present calendar year}) - (\text{year of manufacture}) + 1 \]

Example: During 2003, Peggy’s 1994 Corvette is considered to be 10 years old.

\[ x = (2003 - 1994) + 1 = 9 + 1 = 10 \]

Dependent variable, \(y\): The advertised asking price

1. Discuss why age should be used instead of manufacture year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Asking Price</th>
<th>Year</th>
<th>Asking Price</th>
<th>Year</th>
<th>Asking Price</th>
<th>Year</th>
<th>Asking Price</th>
<th>Year</th>
<th>Asking Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>$17,500</td>
<td>1992</td>
<td>$15,500</td>
<td>1997</td>
<td>$30,000</td>
<td>1995</td>
<td>$24,000</td>
<td>1998</td>
<td>$30,000</td>
</tr>
<tr>
<td>2000</td>
<td>$32,000</td>
<td>1998</td>
<td>$9,500</td>
<td>1993</td>
<td>$17,500</td>
<td>1989</td>
<td>$10,500</td>
<td>1990</td>
<td>$16,000</td>
</tr>
<tr>
<td>1998</td>
<td>$33,000</td>
<td>1993</td>
<td>$16,500</td>
<td>1991</td>
<td>$15,500</td>
<td>1990</td>
<td>$17,500</td>
<td>1996</td>
<td>$21,000</td>
</tr>
<tr>
<td>1994</td>
<td>$16,000</td>
<td>1992</td>
<td>$16,000</td>
<td>2000</td>
<td>$35,000</td>
<td>1990</td>
<td>$14,800</td>
<td>2001</td>
<td>$34,000</td>
</tr>
<tr>
<td>1988</td>
<td>$10,500</td>
<td>2002</td>
<td>$35,000</td>
<td>1993</td>
<td>$14,000</td>
<td>1993</td>
<td>$19,500</td>
<td>1993</td>
<td>$19,500</td>
</tr>
</tbody>
</table>
2. Convert manufacture year to the variable age, \( x \), using the preceding formula.

3. Construct side-by-side vertical dotplots of the asking price for each year of age.

4. Does it appear that the asking price for a Corvette is affected by its age? Describe the effect as pictured on the graph shown for question 3.

5. Complete a one-way ANOVA and test the hypothesis that age has no effect on the asking price. Use \( \alpha = 0.05 \).

6. What effect would using year of manufacture instead of age have on the above analysis?

7. Discuss why using age instead of manufacture year results in a scatter diagram that is more representative of behavior of the price or value of a used car as it ages. Describe how a scatter diagram using year of manufacture as \( x \) would differ from one using age as \( x \).

8. Construct and label a scatter diagram of Peggy’s data.

9. Discuss the relationship between the side-by-side dotplots drawn in question 3 and the scatter diagram drawn in question 8.

10. Determine the equation for the line of best fit.

11. Draw the line of best fit on the scatter diagram.

12. Test the question of the line of best fit to see whether the linear model is appropriate for the data. Use \( \alpha = 0.05 \).

13. Construct a 95% confidence interval for the mean advertised price for 1994 Corvettes.

14. Draw a line segment on the scatter diagram that represents the interval estimate found for question 13.

15. What does the value of the slope, \( b_1 \), represent? Explain.

16. What does the value of the y-intercept, \( b_0 \), represent? Explain.

17. Write a meaningful paragraph answering the general questions of concern:
   a. Is the effect of age on price predictable?
   b. Can a meaningful relationship between the age and the typical asking price for used Corvettes be established?
   c. What is an average asking price for a 1994 Corvette that is for sale?

B Your Own Investigation

Identify a situation of interest to you that can be investigated statistically using bivariate data. (Consult your instructor for specific guidance.)

1. Define the population, the independent variable, the dependent variable, and the purpose for studying these two variables as a regression analysis.

2. Collect 15 to 20 ordered pairs of data.

3. Partition the independent variable values into three or more categories that are meaningful or appropriate for your data.

4. Complete a one-way ANOVA and test the hypothesis that the independent variable has no effect on the dependent variable.

5. Construct and label a scatter diagram of your data.

6. Determine the equation for the line of best fit.

7. Draw the line of best fit on the scatter diagram.

8. Test the equation of the line of best fit to see whether the linear model is appropriate for the data. Use \( \alpha = 0.05 \).

9. Construct a 95% confidence interval for the mean value of the dependent variable at the following value of \( x \): Let \( x \) be equal to one-third the sum of the lowest value of \( x \) in your sample and twice the largest value; that is,

\[
x = \frac{L + 2H}{3}
\]

10. Draw a line segment on the scatter diagram that represents the interval estimate found for question 9.

11. What does the value of the slope, \( b_1 \), represent? Explain.

12. What does the value of the y-intercept, \( b_0 \), represent? Explain.

13. Write a meaningful paragraph comparing and/or contrasting the results from your ANOVA test and the test for the appropriateness of a linear model. What conclusions can you reach based on the results of these tests?