Chapter 1

Discretionary Expenses

1-1 Discretionary and Essential Expenses

1-2 Travel Expenses

1-3 Entertainment Expenses

1-4 Vacation Expenses

1-5 Personal Expenses

CHAPTER OVERVIEW
The course begins with an in-depth study of expenses using the tools of statistics. Students need to understand which statistical tools are appropriate, when to use them, and how to interpret them in the numerous financial contexts in this course. The thread of discretionary and essential expenses in this chapter allows you as the teacher to use student as well as internet researched data to enhance the problem solving situations in need of statistical analysis.

I abhor averages. I like the individual case. A man may have six meals one day and none the next, making an average of three meals per day, but that is not a good way to live.
—Louis Brandeis, Associate Supreme Court Justice

What do you think Brandeis meant in his quote?

As you journey through this course, you examine expenses based on needs and wants. Those expenses will change at different stages of your life. Something that may be an unnecessary purchase for you as a student now could be a necessity when you are a full-time worker. Let’s stop for a minute and compare your financial needs and wants today to your perceived needs and wants 10 years from now. Make two lists, each with the column headings “What do I need?” and “What do I want?” as shown below. In one list, write down as many current expenses you can think of under each heading. In the other, try to imagine what your financial needs and wants may be in the future. Which expenses will change? Which will remain the same? How did you decide which expenses were mandatory and which expenses you had flexibility with?

In this chapter, you use statistics to examine the types of expenses you may have now and in the future. Keeping a mathematical eye on your spending will help you become an informed and skilled consumer.

<table>
<thead>
<tr>
<th>My Wants and Needs Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do I need?</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>My Wants and Needs 10 Years from Now</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do I need?</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
This chapter is all about the way we spend our money. It's about expenses that result from our wants and needs. But what about the way we spend our time? The things we choose to do each day can be broken up into the same two categories as the things we spend our money on—wants and needs. Certainly, there are essential things we absolutely need to spend time on. There are also things we choose to spend our time on based on our wants. Examine these facts on the average minutes per day Americans spend on leisure activities as reported by the United States Bureau of Labor Statistics.

Keep in mind that most people have more leisure time on the weekend than during the week. The time listed in the table is an average amount per day regardless of the day of the week. How does your daily leisure time compare with these averages? In the chapter opener, you were asked to create lists indicating “wants/needs” about money spent by you now and by you in 10 years. Now try making two lists about time spent—one for you now and one for you 10 years from now. Do the money-spent and time-spent lists overlap and interact in any way?

<table>
<thead>
<tr>
<th>Leisure Activity</th>
<th>Average Minutes Spent Per Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watching TV</td>
<td>168.9</td>
</tr>
<tr>
<td>Socializing/communicating</td>
<td>42.4</td>
</tr>
<tr>
<td>Playing games/computer use</td>
<td>26.7</td>
</tr>
<tr>
<td>Reading</td>
<td>19.6</td>
</tr>
<tr>
<td>Sports/exercise/recreation</td>
<td>17.9</td>
</tr>
<tr>
<td>Relaxing/thinking</td>
<td>16.6</td>
</tr>
<tr>
<td>All other leisure activities</td>
<td>24.8</td>
</tr>
</tbody>
</table>

This activity quantifies average time spent on leisure activity. It will be revisited at the end of the chapter. After reviewing the table with students, take the time to have them make the lists suggested in the last paragraph. Students can report their lists to the class and compare their findings.
1-1 Discretionary and Essential Expenses

Objectives
- Identify the difference between essential and discretionary expenses.
- Determine the mean, median, and mode of a data set.
- Use sigma notation to represent and determine the mean of a data set.
- Create and interpret a frequency distribution table.
- Determine the mean, median, and mode of a data set presented in a frequency distribution table.

Key Terms
- gross income
- disposable income
- essential expense
- discretionary expense
- statistics
- data
- measures of central tendency
- mean
- median
- mode
- subscript
- index
- outlier
- skewed data set
- bimodal
- frequency distribution

Warm-Up
Which of the following expressions represents the statement “One half the sum of x and y”?

a. $0.5(x + y)$

b. $\frac{1}{2}(x + y)$

c. $\frac{x + y}{2}$

EXAMINE THE QUESTION
The question posed is the basis for understanding the difference between what is discretionary and what is essential. While it might be easy for students to answer the question in the context of their current lives, it might be more difficult for them to extrapolate to their future selves. Give them some examples of financial needs and wants of adults that you know.

CLASS DISCUSSION
Break the class into three groups: “full-time worker,” “military personnel,” “higher education student.” Have them come up with need/want lists for their assigned category. Share and compare their findings with the class.

What Do You Need? What Do You Want?
In the chapter opener, you were asked to make lists indicating your spending wants and needs now and in the future. When making those lists, you probably considered your income. Right now, you may be on an allowance or have a part-time job. In 10 years, you could be a full-time worker, in the military, or continuing to pursue your education. As your life situation changes, your income, obligations, and expenses will change as well.

Your income before all taxes are deducted is your gross income. Once the taxes are removed from that amount, what remains is known as disposable income. That is, you can dispose of the money (do with it) as you see fit. Disposable income is used for two types of expenses: essential expenses and discretionary expenses. An essential expense is one that can't be eliminated from your day-to-day life. This might include rent or mortgage payments, utility bills, medical expenses, loan payments, and more. Although this type of spending is a necessity, consumers usually look for ways to reduce the amounts. A discretionary expense is a cost for goods or services that are nonessential. This could include movie tickets, coffee, magazine subscriptions, and more.

Wise consumers know how to balance discretionary and essential expenses to live within their means. Often, to do this, consumers use statistics. The branch of mathematics known as statistics involves the mathematical collection, organization, study, interpretation, analysis, and reporting of data. Facts that can be analyzed to obtain more information about a situation are defined as data.
The use of statistics and data can help consumers make responsible decisions to maintain this balance. Data used to represent "typical" values in certain situations so that it is easier to make those decisions are referred to as **measures of central tendency**. There are three measures of central tendency that you have learned in previous math courses and are used in this course:

- **Mean** The "typical" value of a set of scores, is determined by finding the sum of those scores divided by the number of scores.
- **Median** The middle score.
- **Mode** The most frequently occurring item(s).

**Skills and Strategies**

Both essential and discretionary expenses can vary greatly. Here you learn how to use some statistical measures to help you make sense of your expenses.

**EXAMPLE 1**

Alena knows that her morning cup of coffee is most definitely a discretionary expense. She pays $2.75 for a 9-oz cup and was wondering if that is a typical price. On Monday, she asked six of her coworkers what they paid for a 9-oz cup of coffee. Their costs per cup were $2.85, $2.15, $1.95, $3.00, $2.05, and $2.40. How can Alena compare her expense with those of her coworkers?

**SOLUTION** Alena needs to find the mean of the six prices.

\[
\frac{2.85 + 2.15 + 1.95 + 3.00 + 2.05 + 2.40}{6} = 2.40
\]

Notice the fraction bar can be viewed as a grouping symbol that implies parentheses around both the entire numerator and the entire denominator.

The mean is $2.40. Alena's expense is 35 cents above the mean. Based solely on this data, she might want to look around for a less expensive coffee shop in order to save herself some money on this discretionary expense.

Nora is a college student. She needs to make an essential textbook purchase for one of her classes. She researches the cost of the book at her college bookstore, some local booksellers, and some online merchants. The prices per book are $128, $118, $96, $102, $100, $118, $118, and $102. Find the mean of the textbook prices.

**CHECK YOUR UNDERSTANDING**

**Answer** The mean textbook price is $110.25.

**CLASS DISCUSSION**

Students have studied the measures of central tendency since grade school. Ask them to give personal definitions of mean, median, and mode and to identify where they might have used them or seen/heard the terms used before.

**TEACH**

Examples 1–5 offer students opportunities to learn more about the measures of central tendency in the context of spending and expenses. Some of these examples don’t explicitly ask students to determine the mean, median, or mode but rather ask them how they can compare, represent, and interpret data. They might need clarification as to what the question is asking.

**EXAMPLE 1**

The question asks how Alena can compare her expense with those of her coworkers. Before addressing the solution, check that students’ interpretation of the question is accurate.

© 2018 Cengage Learning. All Rights Reserved. This content is not yet final and Cengage Learning does not guarantee this page will contain current material or match the published product.
TEACH
Sigma notation will be new to students. They will probably ask why it is even necessary since they already know how to find an average. Discuss how the notation allows the reader to quickly determine that a sum is involved and which numbers in a list are to be added. Make a connection between what they know about finding the mean to this symbolized notation of it. Before beginning Example 2, offer students a list of expenses that you have researched. Ask them to identify certain elements in the list by using a subscripted variable.

EXAMPLE 2
In this example, students are asked to organize the coffee price data from Example 1 into a list, with each item identified using a subscripted variable. Ask them how they would use sigma notation to symbolize the sum of only the first three prices; the sum of the last four prices; the sum of the second to fifth prices.

CHECK YOUR UNDERSTANDING
Answer \( \bar{x} = \frac{1}{12} \sum x_i = $29 \)

EXTEND YOUR UNDERSTANDING
Answer \( \bar{x} = \frac{1}{6} \sum x_i \)

Check Your Understanding
Addy’s monthly water bills for last year are $27, $31, $30, $26, $25, $27, $37, $33, $32, $28, $26, $26. Express the formula for the mean using sigma notation and calculate the mean water bill for the year.

Extend Your Understanding
Suppose Addy only wanted the mean of the second through the seventh months. Write the formula for this situation in sigma notation.
EXAMPLE 3

Anthony wants to make a discretionary purchase of a basic laptop computer. He checks the prices of a particular make and model listed by seven different vendors on a shopping comparison website. He found these prices: $850, $798, $2,400, $790, $836, $700, $780. He computes the mean as $1,022. This number doesn't seem to be a good representation of the data. How can he find a better representation?

SOLUTION In the prices Anthony found there is an outlier—a piece of data that is extremely different from the rest of the data. When there are outliers, the mean is often not a good representation. In these cases, you can use the median to better represent the data.

When the mean of the data set is not equal to the median, this is a skewed data set. The median is unaffected by the outlier. If the $2,400 price was $924,000, the median would remain the same. The median is resistant to extreme numbers.

There are two methods used to find the median depending on whether there is an even or odd count of numbers. The odd list is addressed here and the even list is in Example 4. When using either list to find the median, arrange the values in ascending order (from least to greatest) or descending order (from greatest to least).

Pair the numbers starting from the ends of the list as shown by the colors. In an odd count list, circle the middle number that remains after the numbers are paired.

The median is the circled number. The number of scores below the median must be the same as the number of scores above the median.

The median is $798. This price is a better representation of the data.

Construct a set of data for a different discretionary expense containing an odd number of scores with the same median as found in Example 3. Identify the type of expense you chose. Explain how the median is the same as the median in Example 3, although the rest of the data are different.
EXAMPLE 4
Explain that in the case with an even number of data pieces, the median may be a number that is not one of the numbers in the original distribution. Remind students that the mean is often a number that is not one of the numbers in the data set.

CHECK YOUR UNDERSTANDING
Answer Answers vary. Sample answer: Cost of a particular computer tablet at six different stores: $700, $750, $800, $834, $980, $1300. Median is $817.

EXAMPLE 5
This problem introduces the mode in a voting context. Students should realize that a voting option can collect the most votes but it might not necessarily collect the majority of the votes.

CHECK YOUR UNDERSTANDING
Answer Answers will vary. Sample answer: Cab fares from Center City to the airport: $35, $42, $48, $56, $60, $41, $56, $66

EXAMPLE 5
A survey was conducted of 880 college students attending the same university. They were offered a list of 10 different Internet service providers and were asked to select the one they prefer. Can a service provider receiving only 89 votes come out on top?

SOLUTION This can be answered by determining the mode. The mode is often used with non-numerical variables, such as in a preference survey. A set can have no mode. If there are two modes, the set is bimodal.

If each of the 880 votes were split evenly among the 10 different Internet service providers, each would get 88 votes and there would be no mode. If one provider received 87 votes, another received 89 votes, and everyone else received 88 votes, the provider with 89 votes could win.

Construct a set of data for a different discretionary expense containing an even number of scores with the same median as found in Example 4. Identify the type of expense you chose. Explain how the median is the same as that in Example 4, although the rest of the data are different.

EXAMPLE 4
Suppose that in Example 3, Anthony had only found the first six laptop prices when he conducted his online search. Determine the median of those prices.

SOLUTION As in Example 3, arrange the numbers in ascending order. Then pair the numbers. Since there are an even number of scores, there is no number left alone in the middle. Circle the last two numbers to be paired.

To find the median, find the mean of the two innermost circled numbers.

Add; then divide by 2.

\[
\frac{798 + 836}{2} = 817
\]

The median is $817. Again, notice that the number of scores below the median is the same as the number of scores above the median, and the median is resistant to extreme scores.

Check Your Understanding

EXAMPLE 4
Explain that in the case with an even number of data pieces, the median may be a number that is not one of the numbers in the original distribution. Remind students that the mean is often a number that is not one of the numbers in the data set.
Frequency Distributions

In your mathematical studies, you have worked with bar graphs, histograms, circle graphs, and line graphs. You will now learn about a table that presents information about central tendency in an easy-to-interpret format. You can use this table when examining data in order to make decisions about essential and discretionary expenses.

EXAMPLE 6

Transportation expenses to and from work are considered essential expenses. Charlie Jane would like to reduce this essential expense by biking to work rather than taking her car. She found 30 different ads both online and in print for the make and model of bicycle she wants to purchase. She made a list of the prices in ascending order.

250, 250, 275, 275, 275, 275, 280, 290, 290, 310, 310, 310, 315, 315, 315, 315, 315, 315, 320, 325, 325, 325, 330, 335, 340, 350, 350, 350

She wants to analyze the prices but is having trouble because there are so many numbers. How can she organize these prices in a helpful format?

SOLUTION Charlie Jane can set up a frequency distribution. A frequency distribution is a table that displays each value and its frequency (the number of times that value appears). The table below lists the price in one column and the number of times that price appeared in her data in the second column.

<table>
<thead>
<tr>
<th>Price, $p$ ($)</th>
<th>Frequency, $f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>2</td>
</tr>
<tr>
<td>275</td>
<td>4</td>
</tr>
<tr>
<td>280</td>
<td>1</td>
</tr>
<tr>
<td>290</td>
<td>2</td>
</tr>
<tr>
<td>310</td>
<td>5</td>
</tr>
<tr>
<td>315</td>
<td>6</td>
</tr>
<tr>
<td>320</td>
<td>1</td>
</tr>
<tr>
<td>325</td>
<td>3</td>
</tr>
<tr>
<td>330</td>
<td>1</td>
</tr>
<tr>
<td>335</td>
<td>1</td>
</tr>
<tr>
<td>340</td>
<td>1</td>
</tr>
<tr>
<td>350</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>30</td>
</tr>
</tbody>
</table>

She adds up the numbers in the frequency column to find the total frequency, which equals the number of pieces of data in her data set. She wants to be sure that she did not accidentally leave out a price. Because there are 30 prices in the set, and the sum of the frequencies is 30, Charlie Jane concludes her frequency distribution is correct.

Use the frequency distribution from Example 6 to find the number of bicycles selling for less than $320.

CHECK YOUR UNDERSTANDING

Answer There are 20 bicycles selling for under $320 (6 + 5 + 2 + 1 + 4 + 2).
EXAMPLE 7
Students may ask why they couldn’t add the individual 30 numbers that they were given in Example 6 and then just divide by 30. They can. Let them know that the product column offers them a more compact way of finding that sum. Make sure that when students add the numbers in the product column, they don’t divide by the number of numbers in that column. They need to divide by 30, not 12. This is a common error.

CHECK YOUR UNDERSTANDING
Answer The mode is $315.

What is the mode of the data set for the frequency table in Example 6?
1. Interpret the quote in terms of what you have learned about essential and discretionary expenses. See margin.

2. For most people, health club membership expenses are considered discretionary. Alli lives in a big city and wants to join a health club. She researched monthly membership costs and found the following for health clubs within a 5-mile radius of her apartment:
   $65, $50, $44, $86, $90, $50, $35, $110, $70, $50, $35, $60, $56
   a. What is the mean monthly membership fee? Round your answer to the nearest cent. $61.62
   b. What is the median monthly membership fee? $56
   c. What is the mode monthly membership fee? $50

3. Kate is a professional musician. She wants to make an essential purchase of an upgraded used bass guitar for her work. She found the following prices for the same make and model bass guitar from various sellers:
   $699, $599, $699, $680, $590, $720, $650, $800
   a. What is the mean price? Round your answer to the nearest cent. $679.63
   b. What is the median price? $689.50
   c. What is the mode price? $699

4. Nick and Liz have decided to move from the city to the suburbs. This means that they will have to make the essential purchase of a car in order to get to work. They researched used 2-year-old cars of the same make, model, condition, and equipped with the same options. They found a website stating that the average price should be $18,500. These are the prices they were quoted:
   $15,500, $18,800, $16,900, $19,900, $18,000, $21,000
   If they continued their search for one more price quote, what would that price have to be so that the mean of all seven of the car prices would be the same as the mean quoted on the website? $19,400

5. Before the last school year began, it was estimated that the average discretionary personal expenses each school year for a student attending a 4-year public college were $2,110. This past summer Ashley decided to poll seven of her friends attending a 4-year public college because she thought that estimate was low. She made a list of their actual school-year expenses:
   $2,800, $1,990, $2,005, $2,400, $1,860, $2,200, $2,000
   a. What is the mean of her friends’ personal expenses? Round your answer to the nearest cent. $2,179.29
   b. How does that average compare with the estimate? about $70 higher
   c. What would Ashley’s actual personal expenses for that school year have to be so that her amount and her friends’ amounts together would have an average of $2,110? $1,625.00
Use the following table to answer questions 6–9.

### Monthly Water Bills

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>$40</td>
<td>$42</td>
<td>$40</td>
<td>$38</td>
<td>$48</td>
<td>$50</td>
<td>$58</td>
<td>$62</td>
<td>$56</td>
<td>$46</td>
<td>$44</td>
<td>$44</td>
</tr>
</tbody>
</table>

6. Write the formula for the mean water bill for the entire year using sigma notation and determine that mean. Round your answer to the nearest cent. See margin.

7. Write the formula for the mean water bill for the first 6 months of the year using sigma notation and determine that mean. Round your answer to the nearest cent. See margin.

8. Write the formula for the mean water bill from April through November using sigma notation and determine that mean. Round your answer to the nearest cent. See margin.

9. Write the sigma notation mean formula for the consecutive 3-month period that would have the highest mean of the year. See margin.

### Daily Cell Phone Minutes Used

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thurs</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
<td>$x_5$</td>
<td>$x_6$</td>
<td>$x_7$</td>
</tr>
<tr>
<td>38</td>
<td>62</td>
<td>40</td>
<td>10</td>
<td>30</td>
<td>55</td>
<td>65</td>
</tr>
</tbody>
</table>

10. Round the following value \(\frac{1}{7} \sum_{i=1}^{7} x_i\) to the nearest minute.

Interpret the answer in the context of the problem. See margin.

11. Round the following value \(\frac{1}{4} \sum_{i=2}^{6} x_i\) to the nearest minute.

Interpret the answer in the context of the problem. See margin.

12. Write the sigma notation mean formula for the 4 consecutive days that would have the lowest mean of the week. See margin.
13. Airline flights can be either discretionary or essential. For most people, the price you pay for where you sit in the plane is a discretionary expense. The seat map for a particular flight is shown here.

The seating options for the color-coded seats are priced as follows:
- First Class: $850
- Deluxe Premium: $540
- Preferred Plus: $400
- Preferred: $320
- Economy: $280

There are a total of 149 seats on this flight. Although seating prices change based on a number of factors, answer the questions below based on the prices listed above.

a. Construct a frequency distribution with column headings “Seat Type,” “Price,” and “Number of Seats.” See margin.

b. If all seats were sold for this flight, what would be the total airline income for the seats? $61,940

c. Determine the mean, median, and mode for seat prices. Round to the nearest cent. Mean: $415.70; Median: $320; Mode: $280

14. There are many cell phone case options on the market. This discretionary item comes in a variety of colors, materials, thicknesses, protection levels, and more. Amit runs a small business that sells computer and phone accessories. He has kept the following inventory of cell phone case sales for similar model phones:

a. Write the formula for the mean in sigma notation and use it to calculate the mean cell phone case price. Round your answer to the nearest cent.

\[ \bar{x} = \frac{\sum x}{n} \]

b. Construct a frequency distribution for the data. See margin.

c. Use the frequency distribution to determine the mean. $40.34

d. Use the frequency distribution to determine the median and the mode. Median: $35.99; Mode: $49.99

15. Medications are essential expenses. DeWitt has composed a price list of antibiotics available at different pharmacies in his neighborhood. In reviewing his list, he can’t find the number of pharmacies selling the antibiotics for $8. Examine the frequency distribution for the prices. Write an expression for the mean.

\[
\bar{y} = \frac{\sum y \cdot f}{\sum f} = \frac{4.10(3) + 4.85(2) + 8x + 12 + 12.5(2)}{3 + 2 + x + 1 + 2} = \frac{59 + 8x}{8 + x}
\]
### Chapter 1 Discretionary Expenses

*Travel is the only thing you buy that makes you richer.*

—Anonymous

#### Objectives
- Determine and interpret cumulative frequency.
- Determine and interpret relative frequency.
- Determine and interpret relative cumulative frequency.
- Model a situation using a spreadsheet.
- Determine and interpret percentiles.

#### Key Terms
- **cumulative frequency**
- **relative frequency**
- **spreadsheet**
- **cell**
- **relative cumulative frequency**
- **percentile**
- **percentile rank**

#### What Will It Cost to Get There?
Examine the following facts reported by the U.S. Department of Transportation and the U.S. Travel Association:

- Over 300 million U.S. residents use some form of transportation each day.
- There are 128.3 million commuters in the United States.
- 3.3 million people travel at least 50 miles to work one-way.
- The net income of all U.S. airline carriers for 2015 was $18,922,412,000.
- Spending by travelers in the United States averages approximately $2.5 billion a day, $104.2 million an hour, $1.7 million a minute, and $28,333 a second.

A travel expense can be either essential or discretionary depending on the situation. Commuting costs for getting to and from work are essential. Airfare for a flight to a beach resort would be discretionary.

The Bureau of Transportation Statistics conducted a survey to gather information about long-distance travel during the summer in the United States. They define “long distance” as a trip of 50 miles or longer and “summer travel” as trips taking place between Memorial Day weekend and Labor Day weekend. In a particular study, they gathered data on over 650,000,000 long-distance summer trips in a single year. Examine this frequency distribution that illustrates the number of trips according to the one-way distance from home.

#### Warm-Up

**a.** What percent of 40 is 8.2? 20.5%

**b.** 66 is 55% of what number? 120

**c.** Write a literal equation that represents $x$% of $z$ is equal to $y$. $y = \left(\frac{x}{100}\right)z$

---

**EXAMINE THE QUESTION**

At the outset, this seems like a very straightforward question. Costs certainly vary depending on the mode of transportation. Impress upon students that there is more to the cost of travel than just the transportation fare.

**CLASS DISCUSSION**

Ask students what costs they might incur when traveling by personal car, rental car, car for hire, bus, train, boat, and plane. Are there any “hidden” costs in these modes of transportation? What ways might they save money when traveling?

**Teach**

The inequality notation used in the “Long-Distance Miles” column of the table appears for the first time here in this course. Ask the students to interpret the numbers and symbols. Pay particular attention to their understanding of whether or not endpoints are included in the intervals.

---

<table>
<thead>
<tr>
<th>Summer Travel</th>
<th>Long Distance Miles (m)</th>
<th># of Long-Distance Summer Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ≤ m &lt; 250</td>
<td>512,460,000</td>
<td></td>
</tr>
<tr>
<td>250 ≤ m &lt; 500</td>
<td>72,270,000</td>
<td></td>
</tr>
<tr>
<td>500 ≤ m &lt; 1000</td>
<td>32,850,000</td>
<td></td>
</tr>
<tr>
<td>$m \geq 1000$</td>
<td>32,850,000</td>
<td></td>
</tr>
</tbody>
</table>
At a glance, you would easily be able to identify the number of long-distance trips for any of the four mileage categories. But what would it take to answer these questions:

- How many of the trips were less than 1000 miles?
- What percentage of the trips were greater than or equal to 500 miles?

While the computation is not difficult to do, this section introduces you to two variations of the frequency table known as the cumulative frequency table and the relative frequency table. Each of these will extend the information given in the basic frequency table that you worked with in Section 1-1, so that questions like those above can be easily answered.

Skills and Strategies

Here you learn how to interpret and construct two new tables as you examine vacation travel and vacation travel expenses.

**EXAMPLE 1**

How many of the trips listed in the Summer Travel table (frequency distribution table) on page 14 were less than 1000 miles?

**SOLUTION** In the Summer Travel table, the column labeled “# of Long-Distance Summer Trips” is the frequency. Those numbers were compiled by sorting the lengths of over 650,000,000 trips into the four mileage categories. To answer this question, you need to add a new category to the table that lists cumulative frequency. This column is created by keeping a running total of all frequencies that are less than or equal to the frequency of that particular interval. The new table is known as a cumulative frequency table and looks like this:

<table>
<thead>
<tr>
<th>Long-Distance Miles (m)</th>
<th># of Long-Distance Summer Trips</th>
<th>Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ≤ m &lt; 250</td>
<td>512,460,000</td>
<td>512,460,000</td>
</tr>
<tr>
<td>250 ≤ m &lt; 500</td>
<td>72,270,000</td>
<td>584,730,000</td>
</tr>
<tr>
<td>500 ≤ m &lt; 1000</td>
<td>32,850,000</td>
<td>617,580,000</td>
</tr>
<tr>
<td>m ≥ 1000</td>
<td>32,850,000</td>
<td>650,430,000</td>
</tr>
</tbody>
</table>

The cumulative frequency column in this table tells the reader the number of trips that are less than or equal to the longest distance in the interval. The number of trips less than 1000 is the sum of the frequencies in the first three rows. This is the cumulative frequency for the interval 500 ≤ m < 1000. There were 617,580,000 summer trips taken that are greater than or equal to 50 miles and less than 1000 miles in length.

How many trips less than 500 miles in length were take
EXAMPLE 2

The Flyt Travel website reported the following airfare booking expenses for the week of July 4th.

What proportion of the tickets sold had a cost greater than or equal to $400 and less than or equal to $499.99?

**SOLUTION**

To answer this question and questions related to it for other intervals, you should create a new column to indicate relative frequency. Relative frequency is the ratio of the frequency of a particular interval to the total number of pieces of data collected. The relative frequency represents the proportion of the data that falls in a particular interval. It is expressed as a decimal rounded to whatever degree of accuracy is needed.

Relative frequency can easily be written as a percent. The relative frequency formulas are given here where $rf$ is the relative frequency, $f$ is the frequency, and $N$ is the total number of frequencies.

Relative frequency expressed as a decimal:

$$rf = \frac{f}{N}$$

Relative frequency expressed as a percent:

$$rf = \frac{f}{N} \times 100$$

To set up the relative frequency table, add a new column. The total number of frequencies represents the total number of bookings reported by Flyt Travel. The sum of the bookings listed in the table is 1,434. Use this to determine the relative frequency.

Ideally, the sum of the relative frequencies should be 1. But, you may get a number very close to 1 because of rounding.

The relative frequency for the $400–499.99$ interval is $0.197$. This represents $19.7\%$ of the total number of bookings reported.

<table>
<thead>
<tr>
<th>Round-Trip Airline Ticket Cost</th>
<th># of Bookings</th>
<th>Relative Frequency (rounded to nearest thousandth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100–199.99$</td>
<td>70</td>
<td>0.049</td>
</tr>
<tr>
<td>$200–299.99$</td>
<td>120</td>
<td>0.084</td>
</tr>
<tr>
<td>$300–399.99$</td>
<td>310</td>
<td>0.216</td>
</tr>
<tr>
<td>$400–499.99$</td>
<td>282</td>
<td>0.197</td>
</tr>
<tr>
<td>$500–599.99$</td>
<td>168</td>
<td>0.117</td>
</tr>
<tr>
<td>$600–699.99$</td>
<td>150</td>
<td>0.105</td>
</tr>
<tr>
<td>$700–799.99$</td>
<td>136</td>
<td>0.095</td>
</tr>
<tr>
<td>$800–899.99$</td>
<td>80</td>
<td>0.056</td>
</tr>
<tr>
<td>$900–999.99$</td>
<td>96</td>
<td>0.067</td>
</tr>
<tr>
<td>$1000 and above</td>
<td>22</td>
<td>0.015</td>
</tr>
<tr>
<td>Total</td>
<td>1,434</td>
<td></td>
</tr>
</tbody>
</table>

**CHECK YOUR UNDERSTANDING**

**Answer** Look at the intervals on the chart. Add the relative frequencies for $700–799.99$ and $800–899.99$: $0.095 + 0.056 = 0.151$. This represents $15.1\%$ of the total bookings.

What percent of the total number of bookings fall within the range of $700–899.99$?
**Spreadsheets**

A spreadsheet is an electronic worksheet that can be used to explore, manipulate, analyze, and interpret data. Spreadsheets allow you to enter data into columns and rows. The intersection of a column and a row is a cell. Cells can contain numbers, words, or formulas. While the structure of a formula may differ based on the software, formulas have a fundamental algebraic basis. In spreadsheet formulas you use an * (asterisk) for the multiplication symbol and a / (forward slash) for the division symbol. You do not use spaces around symbols.

Examine the spreadsheet below. It contains all of the data from the frequency table in Example 2 with an added row labeled “Total.”

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Round-Trip Airline Ticket Cost</strong></td>
<td><strong>#of Bookings</strong></td>
<td><strong>Relative Frequency (rounded to nearest thousandth)</strong></td>
</tr>
<tr>
<td>1</td>
<td><strong>$100–199.99</strong></td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td><strong>$200–299.99</strong></td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td><strong>$300–399.99</strong></td>
<td>310</td>
</tr>
<tr>
<td>4</td>
<td><strong>$400–499.99</strong></td>
<td>282</td>
</tr>
<tr>
<td>5</td>
<td><strong>$500–599.99</strong></td>
<td>168</td>
</tr>
<tr>
<td>6</td>
<td><strong>$600–699.99</strong></td>
<td>150</td>
</tr>
<tr>
<td>7</td>
<td><strong>$700–799.99</strong></td>
<td>136</td>
</tr>
<tr>
<td>8</td>
<td><strong>$800–899.99</strong></td>
<td>80</td>
</tr>
<tr>
<td>9</td>
<td><strong>$900–999.99</strong></td>
<td>96</td>
</tr>
<tr>
<td>10</td>
<td><strong>$1000 and above</strong></td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td><strong>Total</strong></td>
<td>1434</td>
</tr>
</tbody>
</table>

The entries in cells A1 through A12, B1, and C1 are row and column labels. They are not mandatory in a spreadsheet but they help to make sense of the data. It is important to recognize that cells may contain numerical labels such as the ones in column A, but those cells cannot be used for numerical calculations. The frequencies in cells B2 through B11 have been input by hand. All of the other entries have been calculated using spreadsheet formulas. The entry in cell B12 is the sum of the frequencies in cells B2 to B11. A formula can be stored in cell B12 to calculate the sum. If B12 represents the sum, the equation needed is \( B12 = B2 + B3 + B4 + B5 + B6 + B7 + B8 + B9 + B10 + B11 \). You enter the right side of the equation into the cell beginning with the = symbol as shown above.

Spreadsheets have built-in formulas to save time. The long sum formula can be replaced by the formula \( \text{=SUM(B2:B11)} \), where B2 indicates the cell of the first entry in the list, B11 indicates the cell of the last entry in the list, and the colon indicates to add the cell entries in B2 through B11. You can see how this would make the formula entry for a sum a lot easier, especially if you are working with a large data set.
The entries that you see in cells C2 through C11 have not been calculated by hand. For example, a formula has been entered in cell C2 that determines the relative frequency. That formula is \( \frac{B2}{B12} \) (the frequency divided by the total). The formula in cells C3 through C11 are similar to this one. Depending on the program you are using, you can set the number of rounding places for each cell according to the degree of accuracy needed in your calculations. Be aware that the computer retains the entire calculation to many decimal places. In this case, it just shows the value to three decimal places.

**EXAMPLE 3**

Have students identify the cell names of the numbers needed to find the sum. Link the extended sum formula to the compact sum formula. Why might the latter be a better choice? Have students explore the use of parentheses. When are they absolutely needed and when are they not?

**CHECK YOUR UNDERSTANDING**

*Answer* = C2*100 or = B2/B12*100

**EXAMPLE 4**

Extend the relative frequency table on page 16 to determine the percent of the Flyt Travel bookings that are below $600.

**SOLUTION**

The relative frequency column only gives you information for one interval. To answer this question, you need to calculate the cumulative frequency and relative cumulative frequency. The relative cumulative frequency is the ratio of the cumulative frequency of a particular interval to the total number of pieces of data collected. The relative cumulative frequency formulas are given here, where \( rcf \) is the relative cumulative frequency, \( cf \) is the cumulative frequency, and \( N \) is the total number of frequencies.

Relative cumulative frequency expressed as a decimal:

\[
rcf = \frac{cf}{N}
\]

Relative cumulative frequency expressed as a percent:

\[
rcf = \frac{cf}{N} \times 100
\]
Recall the median of a data set. The median is the middle score if the number count is odd and the average of the two middle scores if the number count is even. It divides the data into two equal parts. Since the median is the middle, 50% of the data fall above it and 50% fall below it. Rather than dividing the data into two parts, suppose you divided the data into 100 equal parts. Each part is known as a percentile. Unfortunately, there isn’t complete agreement on the definition of a percentile. Rather than offering and using all of the possibilities, we will stick to the following: A percentile is a statistical measure used to compare data. It indicates what percent of the frequency total appears at or below a particular number. A percentile rank is the percentage of numbers that fall at or below a given number in the list. The median is often referred to as the 50th percentile. Percentile ranks are usually written to the nearest whole percent. Here you learn how to calculate percentiles.

### Percentiles

Examine the entries in the extended table:

<table>
<thead>
<tr>
<th>Round-Trip Airline Ticket Cost</th>
<th># of Bookings</th>
<th>Relative Frequency (rounded to nearest thousandth)</th>
<th>Cumulative Frequency</th>
<th>Relative Cumulative Frequency (rounded to the nearest thousandth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100–199.99</td>
<td>70</td>
<td>0.049</td>
<td>70</td>
<td>0.049</td>
</tr>
<tr>
<td>$200–299.99</td>
<td>120</td>
<td>0.084</td>
<td>190</td>
<td>0.132</td>
</tr>
<tr>
<td>$300–399.99</td>
<td>310</td>
<td>0.216</td>
<td>500</td>
<td>0.349</td>
</tr>
<tr>
<td>$400–499.99</td>
<td>282</td>
<td>0.197</td>
<td>782</td>
<td>0.545</td>
</tr>
<tr>
<td>$500–599.99</td>
<td>168</td>
<td>0.117</td>
<td>950</td>
<td>0.662</td>
</tr>
<tr>
<td>$600–699.99</td>
<td>150</td>
<td>0.105</td>
<td>1100</td>
<td>0.767</td>
</tr>
<tr>
<td>$700–799.99</td>
<td>136</td>
<td>0.095</td>
<td>1236</td>
<td>0.862</td>
</tr>
<tr>
<td>$800–899.99</td>
<td>80</td>
<td>0.056</td>
<td>1316</td>
<td>0.918</td>
</tr>
<tr>
<td>$900–999.99</td>
<td>96</td>
<td>0.067</td>
<td>1412</td>
<td>0.985</td>
</tr>
<tr>
<td>$1000 and above</td>
<td>22</td>
<td>0.015</td>
<td>1434</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Look across the row labeled $500–599.99 to the relative cumulative frequency column. This number represents the proportion of bookings that cost less than $600 (at or below $599.99). The relative cumulative frequency is 0.662. This shows that 66.2% of the bookings cost less than $600.

What does 0.545 in the relative cumulative frequency column represent?

Rounded to the nearest whole number, what percent of bookings fell between $700 and $999.99?

### Percentiles

**CHECK YOUR UNDERSTANDING**

**Answer** 59.4% of the fares were at or below $499.99.

**Check Your Understanding**

**EXTEND YOUR UNDERSTANDING**

**Answer** 100 \( \frac{136 + 80 + 96}{1434} \) = 21.8% rounded to 22%
EXAMPLE 5
It is important that students understand “at or below” when calculating percentiles. If the numbers are arranged in increasing order, this is easy to count. But if numbers are listed at random, encourage students to first make an ordered arrangement and then count. In the table to the right, the calculation of percentiles would be facilitated by a cumulative frequency table.

CHECK YOUR UNDERSTANDING

Answer: \( 100 = \frac{24}{80} = 30 \)

Most cities have some form of taxicab services for short-distance trips. Usually the cab fares are either a flat fee for a specified distance or a metered fee. The metered cabs often start with an initial fee and then charge a fixed rate per fraction of a mile driven. In many cities, wait time and number of stops often affect the fare as well. The website taxifarefinder.com offers a ranking of cities by the cost of a hypothetical taxi trip of a given duration and distance. Of the top 80 most expensive cities, these are the initial cab fees in ranked order:

1.95 2.70 2.85 3.00 3.00 3.00 3.10 3.50 3.50
2.10 2.80 2.85 3.00 3.00 3.00 3.25 3.50 4.00
2.50 2.85 2.85 3.00 3.00 3.00 3.50 3.50 4.00
2.50 2.85 2.85 3.00 3.00 3.00 3.50 3.50 4.00
2.50 2.85 2.85 3.00 3.00 3.00 3.50 3.50 4.00
2.50 2.85 2.85 3.00 3.00 3.00 3.50 3.50 4.95
2.60 2.85 2.85 3.00 3.00 3.00 3.50 3.50 5.75
2.60 2.85 2.85 3.00 3.00 3.00 3.50 3.50 10.00

What is the percentile rank of a $3.00 initial cab fee?

SOLUTION
Determine the number of initial cab fares that are at or below $3.00.
There are 24 initial fares listed that are below $3.00.
There are 32 initial fares listed that are at $3.00.
Therefore, there are 24 + 32, or 56 initial fares listed that are at or below $3.00.
The percentile rank is the percentage of numbers that fall at or below $3.00.
Percentile rank of $3.00 = \frac{56}{80} \times 100 = 70$
The percentile rank of a $3.00 initial cab fee is 70%. This means that 70% of the initial fares listed are at or below $3.00.

Determine the percentile rank for an initial fee of $2.85.
1. Explain how the quote can be interpreted in light of what you have learned in this section. See margin.

2. Winnie needs to travel to Los Angeles. She has decided to go by bus. She researched three different bus companies and compiled the following frequency table for all of the ticket prices available leaving on January 29.

<table>
<thead>
<tr>
<th>Price Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$20–29.99</td>
<td>8</td>
</tr>
<tr>
<td>$30–39.99</td>
<td>10</td>
</tr>
<tr>
<td>$40–49.99</td>
<td>6</td>
</tr>
<tr>
<td>$50–59.99</td>
<td>2</td>
</tr>
<tr>
<td>$60–69.99</td>
<td>9</td>
</tr>
<tr>
<td>$70–79.99</td>
<td>5</td>
</tr>
</tbody>
</table>

   a. Extend the graph by adding a cumulative frequency column. Calculate the six entries for that column and answer the questions below. See Additional Answers.
   b. How many prices are at or below $39.99? 18
   c. How many prices are at or above $40? 22
   d. How many prices are between $29.99 and $70? 27

3. Use the table in Example 2 to answer these questions.

   a. Add a relative frequency column. Calculate the relative frequencies. Round each to the nearest thousandth. See Additional Answers.
   b. Which price range(s) has a relative frequency greater than 0.18 and less than 0.27? $20–29.99, $30–39.99, $60–$69.99
   c. Interpret the relative frequency for the $70–79.99 interval in terms of a percent. 12.5% of the ticket prices were in this price range.

4. Many people travel to the south rim of the Grand Canyon on vacation each year. The National Parks Service keeps records of the number of vehicles entering the park. According to their website, the following table lists the total vehicles that entered the park each month from the beginning of April to the end of September in 2015.

<table>
<thead>
<tr>
<th>2015 Month</th>
<th># of Vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>127,355</td>
</tr>
<tr>
<td>May</td>
<td>140,226</td>
</tr>
<tr>
<td>June</td>
<td>135,000</td>
</tr>
<tr>
<td>July</td>
<td>145,000</td>
</tr>
<tr>
<td>August</td>
<td>175,000</td>
</tr>
<tr>
<td>September</td>
<td>128,850</td>
</tr>
</tbody>
</table>

   a. Add a “Relative Frequency” column to the table and determine all of the entries in that column. Round your entries to the nearest thousandth. See Additional Answers.
   b. Add a “Cumulative Frequency” column to the table and determine all of the entries in that column. See Additional Answers.
   c. Add a “Relative Cumulative Frequency” column to the table and determine all of the entries in that column. Round your entries to the nearest thousandth. See Additional Answers.
   d. What was the average monthly number of vehicles entering the Grand Canyon south rim over this 6-month period? Round up to the nearest whole number. 141,905
   e. What was the median number of vehicles? 137,613
   f. Use your completed chart. What percent of the total number of vehicles over the 6-month period entered in July? 17%
   g. Use your chart. Approximately what percent of all vehicles entering the park did so in April, May, and June combined? Where would you find this information in your chart? 47.3%, the RCF for June.

ANSWERS

1. Travel makes you richer by broadening your horizons and exposing you to new experiences. Making you “richer” does not necessarily mean monetarily richer, rather, it could be educationally, emotionally, or socially richer.
5. Maria and Don Papace are flying to Florida to explore the possibility of moving there. They will be traveling with their two teenage sons. The table below lists information they have researched on the round-trip airfares that are available for purchase on the dates they have chosen to travel.

<table>
<thead>
<tr>
<th>Round-Trip Airfare</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Cumulative Frequency</th>
<th>Relative Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago to Orlando 2 Adults, 2 Children</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$919</td>
<td>4</td>
<td>0.050</td>
<td>4</td>
<td>0.050</td>
</tr>
<tr>
<td>$951</td>
<td>2</td>
<td>0.025</td>
<td>6</td>
<td>0.075</td>
</tr>
<tr>
<td>$999</td>
<td>3</td>
<td>b.</td>
<td>9</td>
<td>0.113</td>
</tr>
<tr>
<td>$1,005</td>
<td>12</td>
<td>0.150</td>
<td>21</td>
<td>0.263</td>
</tr>
<tr>
<td>$1,053</td>
<td>7</td>
<td>0.088</td>
<td>d.</td>
<td>0.350</td>
</tr>
<tr>
<td>$1,079</td>
<td>2</td>
<td>0.025</td>
<td>30</td>
<td>0.375</td>
</tr>
<tr>
<td>$1,133</td>
<td>4</td>
<td>0.050</td>
<td>34</td>
<td>f.</td>
</tr>
<tr>
<td>$1,157</td>
<td>20</td>
<td>c.</td>
<td>54</td>
<td>0.675</td>
</tr>
<tr>
<td>$1,205</td>
<td>5</td>
<td>0.063</td>
<td>59</td>
<td>0.738</td>
</tr>
<tr>
<td>$1,209</td>
<td>7</td>
<td>0.088</td>
<td>e.</td>
<td>0.825</td>
</tr>
<tr>
<td>$1,213</td>
<td>2</td>
<td>0.025</td>
<td>68</td>
<td>0.850</td>
</tr>
<tr>
<td>$1,265</td>
<td>12</td>
<td>0.150</td>
<td>80</td>
<td>g.</td>
</tr>
<tr>
<td>Total</td>
<td>a.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Use the table to determine the missing values a–g. See margin.

6. Four car rental prices were quoted for a 3-day rental. The frequencies are listed. Let \( x \) represent the frequency of the $240 price quote. Use the information shown in the chart to write algebraic expressions for the entries labeled a through d. See margin.

| 3-Day Car Rental | Frequency | Relative Frequency | Cumulative Frequency | Relative Cumulative Frequency |
|------------------|-----------|--------------------|                      |                              |
| $190             | 5         | b.                 |                      |                              |
| $240             | \( x \)   |                    |                      |                              |
| $250             | 3         | c.                 | d.                   |                              |
| $280             | 4         |                    |                      |                              |
| Total            | a.        |                    |                      |                              |

ANSWERS

5a. 80  
5b. 0.038  
5c. 0.250  
5d. 28  
5e. 66  
5f. 0.425  
5g. 1.000

6a. 12 + \( x \)  
6b. \( \frac{5}{12 + x} \)  
6c. \( \frac{8 + x}{12 + x} \)  
6d. \( \frac{8 + x}{12 + x} \)
7. Seaquoia River Cruises offers a 7-day travel package. The prices vary based on the room class booked, as indicated in the following spreadsheet:

<table>
<thead>
<tr>
<th></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>Room</td>
<td>Room Price</td>
<td>Number</td>
<td>Relative</td>
<td>Cumulative</td>
<td>Relative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rooms Frequency</td>
<td>Frequency</td>
<td>Frequency</td>
<td>Cumulative</td>
</tr>
<tr>
<td>2</td>
<td>Standard</td>
<td>$4,750</td>
<td>10</td>
<td>0.1</td>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>Premium Standard</td>
<td>$4,920</td>
<td>10</td>
<td>0.1</td>
<td>20</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>Deluxe</td>
<td>$5,200</td>
<td>15</td>
<td>0.15</td>
<td>35</td>
<td>0.35</td>
</tr>
<tr>
<td>5</td>
<td>Premium Deluxe</td>
<td>$5,850</td>
<td>20</td>
<td>0.2</td>
<td>55</td>
<td>0.55</td>
</tr>
<tr>
<td>6</td>
<td>Deluxe Luxury</td>
<td>$6,962</td>
<td>20</td>
<td>0.2</td>
<td>75</td>
<td>0.75</td>
</tr>
<tr>
<td>7</td>
<td>Premium Luxury</td>
<td>$7,362</td>
<td>15</td>
<td>0.15</td>
<td>90</td>
<td>0.9</td>
</tr>
<tr>
<td>8</td>
<td>Penthouse Balcony</td>
<td>$10,162</td>
<td>10</td>
<td>0.1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Total</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Write the spreadsheet formulas for the indicated cells. See margin.

- a. C9 = SUM(C2:C8)
- b. D3 = C3/C9
- c. D7 = C7/C9
- d. E4 = SUM(C2:C4)
- e. F5 = E5/C9

8. Suppose that two of the Deluxe Luxury rooms in the spreadsheet above were being renovated and unavailable for sale on a particular trip. Therefore, cell C6 would change from 20 to 18. See margin.

- a. Would all, some, or none of the values in column D change based on the change in cell C6? Explain.
- b. Would all, some, or none of the values in column E change based on the change in cell C6? Explain.
- c. Would all, some, or none of the values in column F change based on the change in cell C6?

9. Use the round-trip airfare table in problem 5 above to determine the percentile rank of $1,133. 42.5%

10. Use the cruise ship room price table in problem 7 above to determine the percentile rank of $6,962. 75%
11. Shannon is traveling from New York City to Washington, D.C. She wants to go by train so she can see the views. Since she will be driving home with a family member, she only priced the cost of a one-way ticket on Amtrak for any time of day on February 15. Below is an ordered listing of all fares that were available for selection on that day. See margin.

| $25 | a |
| $38 | b |
| $42 | c |
| $60 | d |
| $65 | e |
| $70 | f |

a. Find the percentile rank for a fare of $119. Interpret your results.
b. Find the percentile rank for a fare of $272. Interpret your results.
c. Based on your answers to parts a and b of this problem, which train fare would have a percentile rank of approximately 82%?

12. For many travelers, the cost of getting to an airport needs to be factored into the travel expense. The table to the rights lists the cost of six different ways to get from the center of a major city to the nearest airport. A researcher polled travelers to determine the number of people who paid each price on a given day. Variables a, b, c, d, e, and f represent the frequencies of the costs. See margin.

a. Write an algebraic expression for the percentile rank of $25.
b. Write an algebraic expression for the percentile rank of $38.
c. Write an algebraic expression for the percentile rank of $42.
d. Write an algebraic expression for the percentile rank of $60.
e. Write an algebraic expression for the percentile rank of $65.
f. Write an algebraic expression for the percentile rank of $70.
How Do Your Entertainment Expenses Vary?

As a teenager, entertainment expenses might comprise a large fraction of your discretionary expenses. Think of all the money Americans spend on entertainment. Think of the inexpensive items. Think of the expensive items. Downloading a song from the Internet can cost less than a dollar, while admission to a sporting event or rock concert for a family can cost several hundred dollars. A diner can sell coffee at a fraction of the price at which a gourmet coffee shop sells it. There truly is a wide range of prices for any and all of our entertainment expenses. Consumers usually need to prioritize and make choices to keep within their budget since entertainment expenses can vary greatly, and they are always discretionary.

Measures of **dispersion** (also referred to as spread) are single numbers that are used to represent the range of the numbers in a given data set. The mean, median, and the mode do not tell the “whole story” of the distribution. A frequency distribution does include how spread out the scores are, but it is sometimes difficult to interpret the dispersion from a table since it has so many entries. There are several single statistics that summarize the dispersion of a distribution of data.
EXAMPLE 2
Maria, a Video Production Club member, noticed the range seemed to “ignore” all the numbers except the highest and lowest. She thought a better measure of dispersion would be to take an average of how far each number is from the mean.

How far, on average, is each entry in the data set from Example 1 from the mean of the distribution?

SOLUTION
The mean is 23. The distance between each number and the mean is called its deviation. The average of how far each number is from the mean is called the mean deviation. Set up a table and subtract the mean from each number.
The mean deviation uses a sum, and it can be represented algebraically.

\[
\text{Mean Deviation} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n}
\]

You can picture each of the deviations from the mean on a number line.

Maria adds the deviations and finds the sum is 0. She divides by 10 (the number of pieces of data) and gets a mean deviation of 0. The mean is located in a central position such that the mean deviation will always be 0 for any distribution. Maria had a good idea, but unfortunately the mean deviation is not a useful measure of dispersion, since it is always equal to 0. Using sigma notation, this fact can be shown algebraically.

\[
\frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n} = 0
\]

Show that the mean deviation for the following data set is 0.

\[8, 12, 34, 55, 68\]

The absolute value of a number is its positive distance from 0. The mean absolute deviation is a statistic that takes the absolute value of each number’s deviation from the mean and averages those absolute values. Find the mean absolute deviation for the data in the Check Your Understanding problem for Example 2.

The mean absolute deviation requires the use of absolute value, which is more cumbersome for mathematicians to make calculations with, so it is not often used as a measure of dispersion. The negative and positive deviations used to compute the mean deviation cancelled each other out. To eliminate these problems, statisticians square each deviation from the mean, and find the average of these squared deviations. The resulting statistic is called the variance. The variance, denoted \(\sigma^2\) (small sigma squared), can be shown algebraically using sigma notation.

\[
\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}
\]

The variance is a very commonly used measure of dispersion.
EXAMPLE 3
While students will eventually use a calculator to find a variance, it is important that they “buy into” the process and why it was created. The middle column of the table reinforces the idea that the mean deviation is always 0.

CHECK YOUR UNDERSTANDING
Answer 63.44.
This was found using the mean of 19.4.

TEACH
Taking the square root is the last step in the logical development of the standard deviation formula that the students have seen in Examples 1–3.

Standard Deviation
Since the variance squares the deviations from the mean, the units of the variance also must be squared. So if a problem’s original units are hours, the units of the variance are hours². If the units are movies, the units of the variance are movies². If the units are dollars, the units of the variance are dollars². These squared measurements don’t make much sense! To measure dispersion using the original units of the problem, statisticians take the square root of the variance to form a measure of dispersion called the standard deviation. The standard deviation is probably the most commonly used measure of dispersion.
EXAMPLE 4
Find the standard deviation for the data from Example 1.

**SOLUTION** In Example 3 the variance for the data from Example 1 was found to be 76.6 movies². To find the standard deviation, just take the square root of the variance.

\[
\sigma = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}
\]

Substitute.

\[
\sqrt{76.6} = 8.75 \text{ movies}
\]

Notice how the units are movies, the original units of the example. You can describe this distribution as having a mean of 23 and a standard deviation of 8.75. You do more interpretation of the standard deviation in Section 1-4.

**EXAMPLE 5**
Pizza is a very popular discretionary expense for high school students. The table below shows the number of times the 80 seniors at Smithtown South High School visited Smithtown Pizza Pals in August. Find the standard deviation. Round to the nearest hundredth.

**SOLUTION** Since there is more than one senior for a given number of visits, you need to take the frequency into account. The mean is 12 visits during August. To find the standard deviation, set up table columns as shown in the next page.

**CHECK YOUR UNDERSTANDING**
Find the standard deviation for the following data set by filling in the columns in the table and using the formula. Round to the nearest hundredth.

\[
\text{Number} \quad (x_i - \overline{x}) \quad (x_i - \overline{x})^2
\]

<table>
<thead>
<tr>
<th>Number</th>
<th>(x_i - \overline{x})</th>
<th>(x_i - \overline{x})^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Add the last column. Divide by 5 to get the average. Then take the square root of the result. The standard deviation is 7.96 to the nearest hundredth.
Number of Visits | Frequency, \( f \) | \( x_i \times f \) | \( x_i - \bar{x} \) | \( (x_i - \bar{x})^2 \) | \( (x_i - \bar{x})^2 \times f \) \\
--- | --- | --- | --- | --- | --- \\
1 | 5 | 12 | 60 | -7 | 49 | 588 \\
2 | 10 | 21 | 210 | -2 | 4 | 84 \\
3 | 12 | 8 | 96 | 0 | 0 | 0 \\
4 | 14 | 18 | 252 | 2 | 4 | 72 \\
5 | 16 | 15 | 240 | 4 | 16 | 240 \\
6 | 17 | 6 | 102 | 5 | 25 | 150 \\
TOTAL | \( n = 80 \) | 960 | | | 1,134 |

First, find the mean using information from the second and third columns.

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{960}{80} = 12
\]

Use the mean to fill in the other columns. Notice that you do not find the sum of the columns that ignore the frequency. Find the standard deviation using the formula.

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{n}} = \sqrt{\frac{1,134}{80}} = 3.76
\]

The standard deviation is 3.76 visits. Your calculator has a menu option that can find the standard deviation, and it will take the frequency column into account. Consult your calculator’s instruction manual.

Donations to charity are discretionary expenses. In a class of 30 students, 6 students donated $5 each to Relay for Life, 15 students donated $10 each, and 9 students donated $20 each. Find the standard deviation for this distribution. Round to the nearest hundredth.

In Example 2 you learned that the mean deviation of any distribution is always 0. Explain why the \( x_i - \bar{x} \) column in the table in Example 5 does not add to 0.
Applications

Nobody goes there anymore—it’s too crowded.
—Yogi Berra

1. Interpret the quote in terms of what you have learned about essential and discretionary entertainment expenses. See margin.

2. The variance of a distribution is 195. What is the standard deviation, rounded to the nearest thousandth? 13.964

3. An electric bill is an essential expense for young people who get their first apartment. The following is a list of Jordan’s monthly electric bills for the past 10 months.

$115, $150, $144, $126, $90, $90, $95, $110, $120, $88

Round your answers to the nearest hundredth.
   a. What is the mean monthly electric bill? $112.80
   b. What is the range? $62
   c. What is the variance? 456.76
   d. What is the standard deviation? $21.37

4. The Vetrone family members are all Cincinnati Reds baseball fans. They went to five games last season. The cost of each game, including parking, tickets, and food, is listed below.

$266, $201, $197, $188, $162

Round your answers to the nearest cent.
   a. What is the mean cost per game attended? $202.80
   b. What is the range? $104
   c. What is the variance? 1,182.96
   d. What is the standard deviation? $34.39
   e. If all five members of the Vetrone family went to each game, what was the mean cost per person to attend each game? $40.56

5. Taking a cruise is a costly discretionary expense. In a recent year, the top five cruise lines in the world had this many passengers:

4,133,000, 2,369,000, 1,295,000, 928,000, 679,000

Round your answers to the nearest integer.
   a. The computations will be easier to work if you view this problem in terms of thousands of passengers. Represent each number in terms of thousands of passengers. 4133, 2369, 1295, 928, 679
   b. What is the mean number of passengers for these five cruise lines? (Give the full number.) 1,880,800
   c. What is the range? (Give the full number.) 3,454,000
   d. What is the standard deviation? (Give the full number.) 1,265,390

6. Janine is analyzing spending habits of students in her high school. She finds a mean monthly discretionary expense of $70 spent on having meals out with friends. The standard deviation of a distribution is 5. What is the variance? 25
ANSWERS

Exercise 7
The customers, since the standard deviation of 2 pounds is relatively more compared to the mean of 6 pounds than compared to a mean of 400 pounds. Although both distributions have a standard deviation of 2, they are not dispersed the same. Standard deviations need to be viewed with respect to their means.

Example 9
Point out to students that if 25 members spend an average of $82 each that they can find the total spent by all 25 members by multiplying 25 and 82. They then can add the 26th person’s ticket price to the total and divide by 26 to compute the new mean.

7. A distribution of newly manufactured amusement park bumper cars has mean 400 pounds and standard deviation 2 pounds. A distribution of customers on line for the bumper-car ride has mean 60 pounds and standard deviation 2 pounds. Which distribution is more spread out? Explain. See margin.

8. Jocelyn compiled data on the number of hours students work per week the summer after senior year. The minimum number of hours in the distribution was 23. The range was 17.
   a. What is the maximum number of hours in the distribution? 40
   b. If the maximum score in the distribution was $M$, and the range is $r$, express the minimum score algebraically. $M - r$

9. An orchestra of 25 members bought tickets, at all different prices, to a concert, and the mean price paid was $82. A new musician joined the orchestra and spent $90 on his ticket to the same concert. What is the new mean price paid when the new musician is included? Round to the nearest cent. $82.31$

10. Ralph wants to join a travel soccer team, a discretionary expense. His parents tell him he must get a job to earn the money to cover the costs. Ralph now views this discretionary expense as an essential expense since he must pay his parents back. The costs of the travel team are listed below.

<table>
<thead>
<tr>
<th>Expense</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>$67.00</td>
</tr>
<tr>
<td>Registration</td>
<td>$50.00</td>
</tr>
<tr>
<td>Tournaments</td>
<td>$775.00</td>
</tr>
<tr>
<td>Travel</td>
<td>$300.00</td>
</tr>
</tbody>
</table>

   a. Find the total cost of being on the team. $1,192
   b. If Ralph takes a summer job and works 10 weeks over the summer, how much must he average each week to cover the travel team expenses? $119.20
   c. If Ralph is paid at a flat rate of exactly $q$ dollars each week, what is the standard deviation of the distribution of his 10 weekly paychecks? 0

11. Sixteen men and sixteen women purchased magazine subscriptions. The distribution of prices had the same mean, minimum score, and maximum score. The red markings on each number line below show where their prices fell.

   a. What is the range of each group? 20
   b. Which group had the larger standard deviation? Explain. See margin.
12. Sometimes outliers can have a significant effect on statistics, and sometimes they don’t. Use the following distribution to determine which statistics are affected by outliers. Round to the nearest hundredth.

\[ 2, 4, 5, 7, 8 \]

a. Find the median of the distribution. 5
b. Find the mean of the distribution. 5.2
c. Find the standard deviation of the distribution. 2.14
d. Make a new distribution consisting of the same first four numbers, but change the 8 to 800. Find the mean, median, and standard deviation. See margin
e. Look at your answers to part d. Which statistic was least changed by the outlier of 800? The median.

13. The mean of a distribution of selected theme park admission prices is $59 with a standard deviation of $3.99. The mean cost of a distribution of a specific home stereo system sold at many different stores is $1,499 with a standard deviation of $4.

a. Which distribution has the lower standard deviation? Theme park admissions since \[ 3.99 < 4. \]
b. Which distribution is more spread out? Explain. See margin

14. A local comic book convention had an admission price of $17. During the week-long show, 19,344 people paid to enter.

a. What is the range of the distribution of admission fees? $0
b. What is the standard deviation of the distribution of admission fees? $0
c. If the range of a distribution is 0, must its standard deviation also be 0? Explain. See margin
d. The number of people who attended each of the 7 days is listed below.

<table>
<thead>
<tr>
<th>Day</th>
<th>Number of People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>178</td>
</tr>
<tr>
<td>Tuesday</td>
<td>246</td>
</tr>
<tr>
<td>Wednesday</td>
<td>222</td>
</tr>
<tr>
<td>Thursday</td>
<td>567</td>
</tr>
<tr>
<td>Friday</td>
<td>1200</td>
</tr>
<tr>
<td>Saturday</td>
<td>5560</td>
</tr>
<tr>
<td>Sunday</td>
<td>11,371</td>
</tr>
</tbody>
</table>

Explain if the range is a good measure of dispersion to describe this distribution of attendance figures. See margin

\[ \text{Example 12d} \]
Mean = 163.6; median = 5; standard deviation = 318.20.

\[ \text{Example 13b} \]
The theme park admissions are more spread out because their standard deviation is greater relative to their mean than the home stereo prices.

\[ \text{Example 14c} \]
Yes. If the range is 0, all the numbers are the same. In that case, all of the mean deviations are 0, and the standard deviation is, as a result, 0.

\[ \text{Example 14d} \]
The range is not a good measure of dispersion since there is an extreme number and no other numbers near it, and the range ignores all the intermediate values.
1-4 Vacation Expenses

For my part, I travel not to go anywhere, but to go. I travel for travel's sake. The great affair is to move.
—Robert Louis Stevenson

Objectives

- Measure dispersion using standard deviation units.
- Compute z-scores.
- Find percentages using the normal curve.
- Compute raw scores using z-scores.

Key Terms

- raw data
- normal curve
- standard score
- z-score
- normal distribution
- bell curve
- asymptotic tails

Warm-Up

Find the $y$-intercept of the straight lines with the following equations.

a. $y = 9x - 11 - 11$

b. $y = 7 + 3x 7$

c. $3y = 6x + 12 4$

How Can Graphs Help Describe Frequency Distributions?

Americans are always on the move. We love to travel, especially in the summer when school is out. In the winter, we love to travel to warmer destinations or go skiing. What are the most popular destinations? How do these popular destinations plan for the correct number of visitors? How much do people spend on souvenirs? Why do hotels change prices during the different travel seasons?

There are many variables travelers and businesses need to take into account to make vacation plans. Consumers need to anticipate costs, and businesses need to anticipate how many visitors they will have. They all can use statistics to aid in their planning.

You might have heard the expression “a picture is worth a thousand words.” In statistics, graphs of distributions are like pictures that describe large quantities of data. Imagine a list of everyone in the United States and their income—the list would be too cumbersome to interpret! Imagine a table of the essential expenses of every household in Florida; think about how measures of central tendency and dispersion could help describe that huge volume of data in a much more user-friendly manner.

You have already learned about percentiles. Percents are commonly used because the raw data—the actual numbers—don’t provide any comparisons. For example, if you hear that 100 students visited the Grand Canyon, you have no frame of reference. If 100 students in an elementary school of 240 students visited the Grand Canyon, that makes it look like that vacation destination is very popular. But if only 100 people in the state of California visited the Grand Canyon, then that would mean the destination is not very popular. Notice how the raw data, without any reference, is tricky to try and interpret. However, if on
a snowy day, 87% of the students in your school slipped and fell on the ice, there is a major danger that must be dealt with. If 1% of the students fell on the ice, the problem is not nearly as severe.

There is a commonly used graph in statistics that uses percents to solve problems. It is prevalent because it accurately portrays the behavior of many real-life data distributions and as a result, it is called the normal curve.

Skills and Strategies

In Sections 1 through 3 you learned about advantages and disadvantages of different measures of dispersion. In the beginning of this section you read how raw data is difficult to interpret because it lacks a frame of reference.

Z-Scores

If a score is 10 away from the mean, and the standard deviation is 2,000, it is relatively close to the mean. If a score is 10 away from the mean and the standard deviation is 2, then the score is very, very far from the mean. So knowing the value of a standard deviation does not tell the whole story. For this reason, we are going to use the standard deviation to convert a raw score into a measure called the standard score—since it is based on the standard deviation. A standard score uses the standard deviation as a unit—it gives the number of standard deviations a number is from the mean. A standard score is usually called a z-score for short. It makes sense to measure raw scores in terms of standard deviations from the mean.

EXAMPLE 1

A summer camp is taking their 220 sixth graders on a trip to an amusement park. For safety purposes, some of the rides have height requirements. The campers' heights have a mean of 56 inches and a standard deviation of 3 inches. What is the z-score for a camper with a height of 62 inches?

SOLUTION The raw score will be represented by \( x \). We are going to use the Greek letter \( \mu \) (mu) to represent the mean just like we used the Greek letter \( \sigma \) to represent standard deviation. (The reason for this is fully explained in Chapter 9). The camper's height of 62 inches is above the mean, and it can be converted to a z-score using the z-score formula:

\[
z = \frac{x - \mu}{\sigma}
\]

Substitute.

\[
z = \frac{62 - 56}{3} = 2
\]

The z-score for the raw score of 62 is 2. This can be interpreted as “62 is 2 standard deviations above the mean.” You can think of the standard deviation as a unit of measurement whenever you work with z-scores.

A camper on the trip has a height of 54 inches. Express his height as a z-score. Round to the nearest hundredth.
EXAMPLE 2

The normal curve is bell-shaped and is often called the bell curve. The middle score represents the mean, and the mean equals the median in a normal distribution. Most of the scores “cluster” around the mean (they are close to the mean). Fewer scores occur as you get farther from the mean. The curve is asymptotic to the horizontal axis. This means that the ends of the curve, called the tails, get closer and closer to the horizontal axis as you move away from the mean. They never touch the horizontal axis.

The Normal Curve

The normal curve is a graph of the normal distribution, and since it reflects so many real-life variables, it is often used in the natural and social sciences to depict the distribution of certain variables. What does the curve look like and what are some of its characteristics?

- The normal curve is bell-shaped and is often called the bell curve.
- The middle score represents the mean, and the mean equals the median in a normal distribution.
- Most of the scores “cluster” around the mean (they are close to the mean).
- Fewer scores occur as you get farther from the mean. The curve is asymptotic to the horizontal axis. This means that the ends of the curve, called the tails, get closer and closer to the horizontal axis as you move away from the mean. They never touch the horizontal axis.

EXAMPLE 2

The height of a certain student on this trip had a z-score of $-0.5$. What is the student’s height in inches?

**SOLUTION** Remember that a z-score of $-0.5$ means the person’s height is half of a standard deviation below the mean. You can use the z-score formula to convert z-scores to raw scores. First, write the z-score formula.

\[
z = \frac{x - \mu}{\sigma}
\]

Substitute.

\[-0.5 = \frac{x - 56}{3}
\]

Multiply both sides by $-3$. 

\[-1.5 = x - 56
\]

Add 56 to both sides.

\[x = 54.5
\]

The z-score of $-0.5$ corresponds to a height of 54.5 inches.

CHECK YOUR UNDERSTANDING

**Answer** $13.8

TEACH

A sketch should be a part of every normal curve problem. Always require students to draw the horizontal axis when sketching normal curves. Students can draw a second horizontal axis underneath and parallel to the first one, and use one axis for raw scores and the other for equivalent z-scores.
• The area under the curve equals 1, which represents 100% of the data in the distribution. Notice that 50% of the data is above the mean and 50% is below on the normal curve.

• The area of any interval under the curve represents the percent of data that is within that interval. To find these areas, you will need to use z-scores and Table 1 on pages 718–719 in the appendix of Financial Algebra.

EXAMPLE 3

Recall the amusement park trip from Examples 1 and 2. A certain ride requires riders to be at least 51 inches tall. The heights are normally distributed with mean 56 and standard deviation 3. Approximately how many of the camp’s 220 sixth graders will not be allowed on the ride?

SOLUTION

You will need to compute z-scores and then use Table 1 on pages 718–719. First draw a sketch of the bell curve. Label where the mean is and where the height of 51 could be. Shade in the area under the curve, below 51 inches.

This area represents the percent of campers who are shorter than 51 inches. Next, convert 51 inches to a z-score.

\[ z = \frac{x - \mu}{\sigma} \]

Substitute.

\[ z = \frac{51 - 56}{3} = -1.67 \]

• Now go to Table 1 and look in the right margin for a z-score of −1.60.
• The right margin of the table gives you the digits for the ones and the tenths place.
• Then, move across the columns to the column headed 0.07. This gives you the digit for the hundredths place.
• Look down the 0.07 column to where it meets the −1.60 row. Copy that number. The number is 0.0475. Therefore, 0.0475 is the area of the shaded portion under the curve.
• Multiply by 100 to convert this to a percent.

\[ 0.0475(100) = 4.75\% \]
Therefore, you can expect about 4.75% of the students to be less than 51 inches tall. Since 4.75% of 220 is 10.45 and you can't interpret a fraction of a student, you can round and say that approximately 10 of the students will not be able to get on the ride.

Use the data from Example 3. If another ride requires riders to be 60 inches tall, what percent of the campers will be able to go on the ride?

Using the data from Example 3, find the percent of students who will be able to go on the ride that requires a height of 51 inches.
EXAMPLE 4

Families of students at Smithtown High School were surveyed about their vacation expenses. The results were normally distributed with mean $2,313 and standard deviation $390. What percent of the families took vacations that cost between $2,000 and $3,000?

**SOLUTION** You will need to convert both expenses to z-scores and look up two areas on the normal curve table. Shading areas under the normal curve will help you plan how to do this.

Label the axis and the mean in each sketch you draw. The shaded area in the first graph is the area you want to compute.

To find this area, find the z-score to the nearest hundredth and use it with the normal curve table to find the area below $2,000.

\[ z = \frac{x - \mu}{\sigma} = \frac{2,000 - 2,313}{390} = -0.80 \]

The area below \( z = -0.80 \) is 0.2119.

Then convert $3,000 to a z-score to the nearest hundredth and use it with the normal curve table to find the area below $3,000.

\[ z = \frac{x - \mu}{\sigma} = \frac{3,000 - 2,313}{390} \approx 1.76 \]

The area below \( z = 1.76 \) is 0.9608.

Subtract the lesser area from the greater area to find the area of the interval from $2,000 to $3,000.

\[ 0.9608 - 0.2119 = 0.7489 \]

This means that 74.89% of the families had vacation expenses between $2,000 and $3,000.

Use the data from the families’ vacation expenses in Example 4 to find the percent of families that had vacation expenses between $1,000 and $2,000. Round to the nearest percent.

**CHECK YOUR UNDERSTANDING**

**Answer** 21%
EXAMPLE 5

A local travel magazine rates hotels using integers from 0 to 100. Last year they rated over 2,000 hotels. The ratings were normally distributed with mean 78 and standard deviation 6.5. How high would a hotel’s rating have to be for it to be considered in the top 10% of rated hotels?

SOLUTION

Shade in the area representing the top 10% under the normal curve. Recall that the score that cuts off the top 10% is also the 90th percentile. Since the normal curve table focuses on the area below a given z-score, to use the table, concentrate on the area of 0.9 below the top 10% cutoff. Look in the body of the normal curve table where the areas are located for an area close to 0.9000.

<table>
<thead>
<tr>
<th>Z*</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.5000</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>.5239</td>
<td>.5279</td>
<td>.5319</td>
</tr>
<tr>
<td>0.1</td>
<td>.5398</td>
<td>.5438</td>
<td>.5478</td>
<td>.5517</td>
<td>.5557</td>
<td>.5596</td>
<td>.5636</td>
<td>.5675</td>
<td>.5714</td>
</tr>
<tr>
<td>0.2</td>
<td>.5793</td>
<td>.5832</td>
<td>.5871</td>
<td>.5910</td>
<td>.5948</td>
<td>.5987</td>
<td>.6026</td>
<td>.6064</td>
<td>.6103</td>
</tr>
<tr>
<td>0.3</td>
<td>.6179</td>
<td>.6217</td>
<td>.6255</td>
<td>.6293</td>
<td>.6331</td>
<td>.6368</td>
<td>.6406</td>
<td>.6443</td>
<td>.6480</td>
</tr>
<tr>
<td>0.4</td>
<td>.6554</td>
<td>.6591</td>
<td>.6628</td>
<td>.6664</td>
<td>.6700</td>
<td>.6736</td>
<td>.6772</td>
<td>.6808</td>
<td>.6844</td>
</tr>
<tr>
<td>0.5</td>
<td>.6915</td>
<td>.6950</td>
<td>.6985</td>
<td>.7019</td>
<td>.7054</td>
<td>.7088</td>
<td>.7123</td>
<td>.7157</td>
<td>.7190</td>
</tr>
<tr>
<td>0.6</td>
<td>.7257</td>
<td>.7291</td>
<td>.7324</td>
<td>.7357</td>
<td>.7389</td>
<td>.7422</td>
<td>.7454</td>
<td>.7486</td>
<td>.7517</td>
</tr>
<tr>
<td>0.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
<td>.7673</td>
<td>.7704</td>
<td>.7734</td>
<td>.7764</td>
<td>.7794</td>
<td>.7823</td>
</tr>
<tr>
<td>0.8</td>
<td>.7881</td>
<td>.7910</td>
<td>.7939</td>
<td>.7967</td>
<td>.7995</td>
<td>.8023</td>
<td>.8051</td>
<td>.8078</td>
<td>.8106</td>
</tr>
<tr>
<td>0.9</td>
<td>.8159</td>
<td>.8186</td>
<td>.8212</td>
<td>.8238</td>
<td>.8264</td>
<td>.8289</td>
<td>.8315</td>
<td>.8340</td>
<td>.8365</td>
</tr>
<tr>
<td>1.0</td>
<td>.8413</td>
<td>.8438</td>
<td>.8461</td>
<td>.8485</td>
<td>.8508</td>
<td>.8531</td>
<td>.8554</td>
<td>.8577</td>
<td>.8599</td>
</tr>
<tr>
<td>1.1</td>
<td>.8613</td>
<td>.8665</td>
<td>.8686</td>
<td>.8708</td>
<td>.8729</td>
<td>.8749</td>
<td>.8770</td>
<td>.8790</td>
<td>.8810</td>
</tr>
<tr>
<td>1.2</td>
<td>.8849</td>
<td>.8869</td>
<td>.8888</td>
<td>.8907</td>
<td>.8925</td>
<td>.8944</td>
<td>.8962</td>
<td>.8980</td>
<td>.8997</td>
</tr>
</tbody>
</table>

The closest area is 0.8997. This area corresponds to a z-score of 1.28. Change this z-score to a raw score.

\[ z = \frac{x - \mu}{\sigma} \]

Substitute.

\[ 1.28 = \frac{x - 78}{6.5} \]

Multiply both sides by 6.5.

\[ x - 78 = 8.32 \]

Add 78 to both sides.

\[ x = 86.32 \]

A score of 87 will allow the hotel to advertise that it is in the top 10%.

Using the travel magazine statistics from Example 5, find out what a hotel’s rating would have to be below to be ranked in the lower quartile of hotels. Round to the nearest integer.
1. Interpret the quote in terms of what you have learned about essential and discretionary vacation expenses.

2. A recent survey by the American Automobile Association showed that a family of two adults and two children on vacation in the United States will pay an average of $247 per day for food and lodging with a standard deviation of $60 per day. Assuming the data are normally distributed, find, to the nearest hundredth, the z-scores for each of the following vacation expense amounts.
   a. $197 per day. -0.83
   b. $277 per day. 0.50
   c. $310 per day. 1.05

3. If a family from the data set in question 2 had a z-score of 1.5, what was their daily expense for food and lodging? $337

4. Use the data from question 2. If the data are normally distributed, find the percentage of families who spent:
   a. Less than $167 per day. 9.18%
   b. Less than $367 per day. 97.72%
   c. More than $247 per day. 50%
   d. More than $350 per day. 4.27%
   e. Less than $67 per day. 0.13%
   f. Between $200 and $300 per day. 59.29%
   g. Between $360 and $400 per day. 2.47%
   h. More than the median. 50%
   i. Less than the mean. 50%

5. Using the data from question 2, you can compute the mean food and lodging cost of a 10-day vacation by multiplying the per-day mean by 10. The standard deviation for the 10-day vacation can be found by multiplying by 10 also.
   a. What is the mean for food and lodging for a 10-day vacation? $2,470
   b. What is the standard deviation of the distribution for the 10-day vacation? $600
   c. What is the variance of the distribution for the 10-day vacation? 360,000

6. A recent Internet survey showed per-day lodging expenses at Disney World in Orlando range from less than $100 for a campsite to more than $2,000 for a top luxury room. The cost-per-day average of the top 10 most popular hotels at Disney World is $348.
   a. If the standard deviation is $80, find the z-score for a per-day expense of $400. 0.65
   b. If the data are approximately normally distributed with mean $348 and standard deviation $80, what percent of the hotels have per-day expenses greater than $268? 84.13%
7. International travel is usually more expensive than domestic travel. A recent survey found that the average per-person cost of a 12-day international vacation is $1,755. This includes transportation, food, lodging, and entertainment.
   a. If the data are normally distributed with standard deviation $390, find the percentage of vacationers who spent less than $1,200 per day. Round to the nearest hundredth of a percent. \(7.78\%\)
   b. Find the per-day expense for one of these travelers who had a z-score of \(-2.1\). $936
   c. If the data are normally distributed with standard deviation $490 instead of $390, would the answer to question 7b become lower or higher? Lower

8. A Caribbean resort has a nightly limbo contest on the beach. Participants must be less than 64 inches tall. The distribution of heights of adult American men is approximately normal with mean 69 inches and standard deviation 2.5 inches.
   a. What percent of adult American males could enter this contest? 2.28%
   b. How tall would a man have to be for his height to be in the top 15%? 71.6"

9. Visitors to the Royal Ranch Golf Paradise shoot an average of 88 on their 18-hole golf course. The standard deviation is 12. Visitors to the Glen Oaks Golf Resort shoot an average of 97 with a standard deviation of 14. (In golf, the lower score is the better score.) Lisa was vacationing at Royal Ranch and shot an 89. Angela was golfing at Glen Oaks and shot a 90. Explain who is the better golfer based on these results.

10. Getaway Sports is a vacation destination for the serious sports fanatic. Rajesh is in charge of equipment for the complex. The last shipment of ping-pong balls was normally distributed with a mean weight of 2.7 grams and a standard deviation of 0.1 gram. The 100-pound weights he orders from the manufacturer for the fitness center are normally distributed with mean weight 100.2 pounds and standard deviation 0.1 pound.
    a. Which distribution has the larger standard deviation? 100-pound weights
    b. Which distribution is more dispersed? Explain. See margin.

11. A finance magazine did a survey and found that the average American family spends $1,600 on a summer vacation. If the distribution is normal with standard deviation $411, find the amount a family would have spent to be the 80th percentile.
1-5 Personal Expenses

If you want to understand today, you have to search yesterday.
—Pearl S. Buck, American writer and novelist

Objectives
- Graph bivariate data.
- Interpret patterns and trends based on data.
- Construct a scatter plot.
- Fit a linear regression equation to a scatter plot.
- Determine the linear regression equation.
- Find and interpret the correlation coefficient.
- Make predictions based on lines of best fit.

Key Terms
- univariate data
- bivariate data
- scatter plot
- trend
- correlation
- causal relationship
- explanatory variable
- response variable
- lurking variable
- linear regression analysis
- linear regression equation
- independent variable
- dependent variable
- domain
- interpolation
- extrapolation
- correlation coefficient

Warm-Up
a. Graph the line represented by the equation $y = -3 + 2x$

b. Find the slope and the $y$-intercept of the line $70 = -2y + 8x$

See Additional Answers.

How Can Past Trends Predict Future Occurrences?
Examine this bar graph that depicts how teenagers spend their money.

EXAMINE THE QUESTION
In order to understand the question, students need to understand the meaning of the words “trends,” “predict,” and “occurrences.” A trend is a general direction in which something is heading. To predict is to estimate or forecast something that will happen in the future based on something that has happened in the past or is happening now. An occurrence is an event. This question sets the stage for using statistical prediction tools in this and future chapters.
Compare your spending habits with those depicted in the graph. Where are they the same? Where are they different? Think about your personal expenses. Are any categories missing from this list? In this section, you learn how to find patterns in data about personal spending. You do this both graphically and algebraically. The data we have been working with in this chapter has all been single sets of numbers known as univariate data. For example, when a person keeps a list of weekly expenses, the data in the list are univariate data. Using univariate data, you can find measures of central tendency and measures of spread. But sometimes data are collected and presented in pairs. Data represented by pairs of numbers that possibly show a relationship between the paired numbers is called bivariate data. If you keep a list of the days of the month and the amount of disposable money left from your monthly allowance, the set is bivariate data.

A scatter plot is a graph that shows bivariate data using points on a graph. Scatter plots may show a general pattern, or trend, within the data. A trend is a relationship between two variables. A trend may show a correlation, or association, between two variables. A positive correlation exists if the value of one variable increases when the value of the other variable increases. A negative correlation exists if the value of one variable decreases when the value of the other variable increases.

A trend may also show a causal relationship, which means one variable caused a change in the other variable. The variable that causes the change in the other variable is the explanatory variable. The affected variable is the response variable. While a trend may indicate a correlation or a causal relationship, it does not have to. If two variables are correlated, it does not mean that one caused the other.
You can graph a scatter plot by hand, by using graphing software, or by using a graphing calculator.

**EXAMPLE 1**

Ravi is a 16-year-old high school student who has a part-time job after school. He is looking to see if there are any trends that relate monthly income to monthly clothing purchases among his high school peers. He thinks these two variables might be related and wants to investigate them using a scatter plot. Below is a list of ordered pairs \((x, y)\) where \(x\) represents the monthly income and \(y\) represents monthly clothing expenses for 10 of his classmates.

\[
(205, 36), (242, 45), (268, 57), (296, 63), (303, 69),
(327, 64), (339, 75), (344, 80), (355, 82), (380, 84)
\]

Construct a scatter plot on graph paper by hand. Then enter the data into a graphing calculator or use graphing software to create a digital scatter plot.

**SOLUTION** In each ordered pair, the first number is the monthly income in dollars and the second number is the average monthly clothing expense in dollars. The scatter plot is drawn by plotting the points with the given coordinates. Choose a scale for each axis that allows the scatter plot to fit in the required space. To choose the scale, look at the greatest and the least numbers that must be plotted for each variable. Label the axes accordingly. Then plot each point. Notice that you do not connect the dots in a scatter plot.

Use the statistics features on your graphing calculator or software to graph the scatter plot. Your display should look similar to this.

In both graphs, you can see that a positive correlation appears to exist between monthly income and average monthly clothing expense since as one variable increases, so too does the other variable.

Use the scatter plot above. If a student has a monthly income of about $250, how much do you predict he might spend on clothing in a month? Explain.
EXAMPLE 2

For many people, the cost of contact lenses is a discretionary expense. Jamill buys contact lenses through an online retailer. He has found the type he needs is sold six lenses to a box. The retailer has a sliding price based on the number of boxes purchased according to the following table:

<table>
<thead>
<tr>
<th>Boxes</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>20</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price per Lens</td>
<td>$6.00</td>
<td>$5.90</td>
<td>$5.40</td>
<td>$5.20</td>
<td>$4.80</td>
<td>$4.10</td>
<td>$3.80</td>
<td>$3.10</td>
<td>$2.50</td>
<td>$1.70</td>
</tr>
</tbody>
</table>

Examine the scatter plot created using the bivariate data in the table where the $x$-value represents the number of boxes purchased and the $y$-value represents the price per lens for that purchase.

What do you notice about the number of boxes purchased and the price per lens? Is there an explanatory variable and a response variable?

SOLUTION

Jamill noticed the higher the quantity of boxes ordered, the lower the price per lens charged. Because $y$-values decrease as $x$-values increase, the correlation is negative. Additionally, the increase in boxes purchased caused the drop in the price per lens. This could have occurred for a number of reasons such as the retailer being willing to take less of a profit for an increase in sales. Therefore, boxes sold is the explanatory variable and price per lens is the response variable.

The Young Entrepreneurs Club at North Lane High School has been given permission to sell cold water bottles and hot chocolate to students as they walk into school each day as part of a year-long fundraising event. For the entire school year, the students kept track of the outside temperature and the sales. Two scatter plots were graphed, one for each drink, with temperature on the horizontal axis. Which scatter plot do you think showed a positive correlation? What was the causation? Explain.
EXAMPLE 3
Tiana was looking at the teen spending chart on page 43. She noticed that the percentage teens spend on personal care/cosmetics/accessories is the same as the percentage teens spend on shoes. She decided to poll all of the girls on both the junior varsity and varsity lacrosse teams at her school. She created the scatter plot on the right with the school year amount spent on personal care/cosmetics/accessories and the school year amount spent on shoes as the variables.

What conclusions might you draw from the scatter plot? Does it show that the amount spent on personal care/cosmetics/accessories causes the amount spent on shoes?

SOLUTION The scatter plot shows a positive correlation. The increased amount spent on personal care/cosmetics/accessories does not cause an increase in the amount spent on shoes. Most likely, both amounts increase with student age and student income. Notice the variables of age and income are not mentioned in the problem statement and cannot be classified as explanatory or response variables. In this case, age and income would be called lurking variables. A lurking variable is an external variable that may influence how variables are perceived to be related to one another. Because of the unseen lurking variable, people tend to make causality assumptions between variables where they do not exist. Keep in mind that if two variables are correlated, they are associated in some way but one does not necessarily have to cause the other. The amount spent on cosmetics does not cause the amount spent on shoes to be a certain value.

Think of an example of a negative correlation for which there is no causal relationship.

EXAMPLE 4
The scatter plot shows the relationship between the number of coffee purchases made during a semester and the grade point average for that semester for 10 students. Describe any patterns you see in the data.

SOLUTION As the number of coffee purchases increases, the GPA does not increase, so there is no positive correlation. As the number of coffee purchases increased, the GPA did not decrease, so there is no negative correlation. There is no trend in the data, so there is no correlation.

High school students collected data on the price of a cell phone purchase and the tax on that purchase. The students created a table of ordered pairs for these two amounts and plan to construct a scatter plot. Should there be a positive correlation, a negative correlation, or no correlation?
Linear Regression Analysis

Many scatter plot points can be approximated by a single line that best fits the scattered points. This approximation is made by using linear regression analysis. In linear regression analysis, the equation of a line is found that best fits the points in the scatter plot. This equation may be called a linear regression equation or a line of best fit. The equation is of the form \( y = mx + b \), where \( m \) is the slope and \( b \) is the \( y \)-intercept. Both \( m \) and \( b \) are fixed numbers that are needed to graph the line. A linear equation shows a relationship between two variables, usually \( x \) and \( y \). In general, one is the independent variable and the other is the dependent variable. An independent variable can take on any permissible value. It does not depend on the values of any other variable. The independent variable can be considered the “input” for an equation. The dependent variable is calculated based on the value of the independent variable. It is the “output” for a given input. The value of the dependent variable directly depends on the value of the independent variable.

Both the independent and dependent variables in a linear equation are used to calculate the slope. Slope is interpreted as a change in \( y \) over a change in \( x \). The Greek letter delta, \( \Delta \), is used to represent change. The formula for slope is \( m = \frac{\Delta y}{\Delta x} \).

The \( y \)-intercept \( b \) is the value on the \( y \)-axis where the line crosses that axis. This regression line can be used to display a trend and predict corresponding variables for different situations. It is more efficient to rely on the single line rather than the scatter-plot points because the line can be represented by an equation.

Recall from a previous math course that the domain is a set of first elements of the ordered pairs, and the range is the set of corresponding second elements. Given an \( x \)-value within the domain you can predict corresponding \( y \)-values by interpolation. Given an \( x \)-value outside of the domain you can predict corresponding \( y \)-values by extrapolation.

The scatter plot below is from Example 1 at the beginning of this section in which Ravi examined monthly income and clothing expenses. The linear regression equation can be determined for the line that best fits the scatterplot points.

The horizontal and vertical labels are included to identify the axes, but will not be shown on a calculator display. Generally, the distance the points lie from the line of best fit determines how good a predictor the line is. If most of the points lie close to the line, the line is a better predictor of the trend of the data than if the points lie far from the line. If the points lie far from the line, the line is not good for predicting a trend.

The correlation coefficient, \( r \), is a number between \(-1\) and \(1\) inclusive that is used to judge how closely the line fits the data. Negative correlation coefficients show negative correlations, and positive correlation coefficients show positive correlations. If the correlation coefficient is near 0, there is little or no correlation. Correlation coefficients with an absolute value greater than 0.75 are strong correlations. Correlation coefficients with an absolute value less than 0.3 are weak correlations. Any other correlation is a moderate correlation.
EXAMPLE 5

Find the linear regression equation for Ravi's scatter plot in Example 1. Round the slope and the $y$-intercept to the nearest hundredth. The points are given here:

$$(205, 36), (242, 45), (268, 57), (296, 63), (303, 69)$$
$$(327, 64), (339, 75), (344, 80), (355, 82), (380, 84)$$

**SOLUTION** Although it is possible to find the linear regression equation using paper and pencil, it is a lengthy process. Using the linear regression feature on a graphing calculator, graphing software, or in a spreadsheet produces more accurate results.

Enter the ordered pairs. Then use the statistics menu to calculate the linear regression equation. Rounding the slope and the $y$-intercept to the nearest hundredth, the equation of the regression line is $y = 0.29x - 22.02$.

Note that graphing programs may differ in the letters they use to represent the slope and the $y$-intercept. Remember the coefficient of the independent variable, $x$, is the slope.

Find the linear regression equation for the scatter plot defined by these points: $(1,56), (2,45), (4,20), (3,30), \text{ and } (5,9)$.

EXAMPLE 6

Interpret the slope as a rate for Ravi's linear regression equation. Use the equation from Example 5.

**SOLUTION** The $x$-values represent monthly income and the $y$-values represent the average monthly clothing expense. The slope is interpreted as the rate of clothing expense dollars per income dollars. The slope is 0.29 and can be represented as the rate $\frac{0.29}{1}$. This ratio means that for every 1 dollar increase in income, there is a 29 cent increase in clothing expense.

Approximately how much of an increase in clothing expense is there if income increases by $2$?

If you use the regression line graph on page 48 to determine that a monthly income of $280 would approximately correlate with a clothing expense of $59, would you be using interpolation or extrapolation? Explain.
EXAMPLE 7
Before reading the solution, ask students if this is an example of interpolation or extrapolation and why.

CHECK YOUR UNDERSTANDING
Answer $0.58.
Students can use the equation to show the $0.58 increase between two incomes that are $2 apart.

Example 7
Ravi wants to determine an appropriate monthly amount to budget for his own clothing based on the analysis of the data in Examples 1 and 5. He works two part-time jobs and gets a small allowance from his parents. His total income is $420 per month. What is a reasonable amount for his clothing expense?

SOLUTION The linear regression equation tells Ravi the approximate monthly clothing expense given a specific monthly income amount. Substitute $420 for x in the equation, and compute y, the clothing expense he should budget for.

Equation of the regression line.

\[ y = 0.29x - 22.02 \]

Substitute 420 for x.

\[ y = 0.29(420) - 22.02 \]

Simplify.

\[ y = 99.78 \]

Based on the data he collected, it would be reasonable to assume an expense of about $100 for clothing per month. This is an example of extrapolation because $420 was not between the low and high x values of the original domain.

Use the equation above to determine approximately how much of an increase in clothing expense there is if income increases by $2?

Example 8
Determine the correlation coefficient to the nearest hundredth for the relationship in for Ravi’s data. Interpret the correlation coefficient.

SOLUTION Use your graphing device, software, or website to find the correlation coefficient.

Round r to the nearest hundredth.

\[ r = 0.98 \]

Because 0.98 is positive and greater than 0.75, there is a strong correlation between the monthly income and the amount spent on clothing.

Find the correlation coefficient to the nearest thousandth for the data in Check Your Understanding for Example 5. Interpret the correlation coefficient.

Carlos did analyses of two different data sets. One had a correlation coefficient of −0.28. The other had a correlation coefficient of 0.28. Explain the similarities and the differences between these two r values.
1. Interpret the quote in terms of what you have learned about essential and discretionary expenses. See margin.

2. You get paid on the first day of each month. You cash your check and pay all of your essential expenses. You keep the balance in your “discretionary spending” envelope. A scatter plot shows the number of days that have passed since you were paid and the amount left in your discretionary spending envelope. The explanatory $x$-variable is the number of days that have passed. The response $y$-variable is the amount left in your envelope. Is there a positive or negative correlation? Explain. See margin.

3. Examine each scatter plot. Identify each as showing a positive correlation, a negative correlation, or no correlation.

4. The MyTunes song app sells music downloads. Over the past few years, the service has lowered its prices. The table shows the price per song and the number of songs downloaded per day at that price.

   a. Examine the data without drawing a scatter plot. Describe any trends you see. See margin.
   b. Draw a scatter plot. Describe the correlation. See margin.
   c. Approximate the number of downloads at a price of $0.54 per song. Explain your reasoning. See margin.

   **Price per Song** | **Number of Downloads (in thousands)**
   --- | ---
   $0.89 | 1,212
   $0.79 | 1,704
   $0.69 | 1,760
   $0.59 | 1,877
   $0.49 | 1,944
   $0.39 | 2,011

4a. As the price increases, the number of downloads decreases.
4b. Negative correlation.
4c. Approximately 1,911 downloads. $0.54 is the average of $0.49 and $0.59. So the average of 1,944 and 1,877 yields a good approximation.
Chapter 1 Discretionary Expenses

5. At the end of 2015, the U.S. Bureau of Economic Analysis posted the following data on the U.S. disposable personal income by month for the first 11 months of that year as indicated in this table. See margin.

<table>
<thead>
<tr>
<th>Month</th>
<th>Disposable Personal Income (billions of U.S. dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13265.6</td>
</tr>
<tr>
<td>2</td>
<td>13315.8</td>
</tr>
<tr>
<td>3</td>
<td>13304.6</td>
</tr>
<tr>
<td>4</td>
<td>13352.4</td>
</tr>
<tr>
<td>5</td>
<td>13429.3</td>
</tr>
<tr>
<td>6</td>
<td>13352.5</td>
</tr>
<tr>
<td>7</td>
<td>13407.8</td>
</tr>
<tr>
<td>8</td>
<td>13456</td>
</tr>
<tr>
<td>9</td>
<td>13477.1</td>
</tr>
<tr>
<td>10</td>
<td>13601.9</td>
</tr>
<tr>
<td>11</td>
<td>13630.9</td>
</tr>
</tbody>
</table>

5a. For the most part, as the months increased, the disposable income increased.

5b. Positive correlation.

5c. The plot shows an increase in the disposable income. The income in December might be $13,640B.

6. Back-to-school shopping is a very big business. Some may consider these expenses essential while others look at them as discretionary costs. However you look at it, the data of the average dollars spent per family with school-age children is very interesting. Examine the chart below with information from the National Retail Federation. See margin.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-to-School Spending in Dollars Per Family</td>
<td>$483.28</td>
<td>$443.77</td>
<td>$527.08</td>
<td>$563.49</td>
<td>$594.24</td>
<td>$548.72</td>
<td>$606.40</td>
<td>$603.63</td>
<td>$688.82</td>
<td>$634.79</td>
<td>$669.28</td>
<td>$630.36</td>
</tr>
</tbody>
</table>

6a. Overall, there was an increase in spending but some individual years showed a decrease.

6b. There is a positive correlation.

7. Over the past 4 years, Reggie noticed that as the price of a slice of pizza increased, her college tuition also increased. She found the correlation coefficient was $r = 0.49$. See margin.

7a. Graph A shows a moderate correlation, which could be $r = 0.49$. The points on graph B appear to lie close to a straight line. This implies a strong correlation.

7b. Neither since there is no causal relationship.

TEACH Data Entry

Students need to double-check calculator list entries before making calculations. The calculator will display incorrect answers just as quickly as it will correct answers if data are entered incorrectly. It is very easy to make a mistake entering data. Also remind them to check the number of pieces of data they entered to make sure it agrees with the original list.
8. Describe each of the following correlation coefficients using the terms strong, moderate, or weak and negative, or positive.
   a. $r = 0.21$ \textit{weak pos.}  
   b. $r = -0.87$ \textit{strong neg.}  
   c. $r = 0.55$ \textit{moderate pos.}  
   d. $r = -0.099$ \textit{weak neg.}  
   e. $r = 0.99$ \textit{strong pos.}  
   f. $r = -0.49$ \textit{moderate neg.}

9. Which of the following scatter plots shows a correct line of best fit? c.

10. Is it possible for a linear regression line to go through every point on a scatter plot? Is it possible for a linear regression line to not go through any point on a scatter plot? Explain.

11. Use these data from Example 2 about the number of boxes purchased and the price per contact lens to answer a through g below.
   (1, 6.00), (2, 5.90), (4, 5.40), (8, 5.20), (10, 4.80), (12, 4.10), (14, 3.80), (16, 3.10), (20, 2.50), (24, 1.70)
   a. Find the linear regression equation. Round the slope and the $y$-intercept to the nearest hundredth. $y = -0.19x + 6.36$
   b. What is the slope of the linear regression line? $-0.19$
   c. What are the units of the slope when it is expressed as a rate? \textit{Lens price per box}
   d. Based on the linear regression equation, what would be the price per lens if he bought 22 boxes? Round to the nearest cent. $\$2.18$
   e. Is your answer to part d interpolation or extrapolation? Explain. \textit{Interpolation; 22 is in the domain.}
   f. Find the correlation coefficient to the nearest hundredth. $r = -0.99$
   g. Describe the correlation. \textit{Strong negative}

12. The table gives the number of songs purchased from MyTunes at different prices per song.
   a. Find the equation of the linear regression line. Round the slope and the $y$-intercept to the nearest hundredth. $y = -1380.57x + 2634.90$
   b. What is the slope of the linear regression line? $-1380.57$
   c. What are the units of the slope when it is expressed as a rate? \textit{Downloads per dollar}
   d. Based on the linear regression line, how many thousands of downloads would MyTunes expect if the price were changed to $0.45? Round to the nearest integer. \textit{2014}
   e. Is your answer to part d interpolation or extrapolation? Explain. \textit{Interpolation; \$0.45 is in the domain.}
   f. Find the correlation coefficient to the nearest hundredth. $r = -0.90$
   g. Describe the correlation. \textit{Strong negative}
13. Examine the graph below. It shows the global iPhone phone sales by year in millions of units sold.

a. Write nine ordered pairs with the first coordinate being the year number and the second being the units sold in millions as shown here. \(1,400,000\) units sold in 2007 would be represented by \((1, 1.4)\) since 2007 is year 1 of the data listed. To save data entry time, you can use 1.4 rather than \(1,400,000\) as long as you remember that your \(y\)-coordinates will always be in millions when using and reporting data and findings. See margin.

b. Construct a scatter plot. See margin.

c. Find the equation of the regression line. Round the slope and the \(y\)-intercept to the nearest hundredth. \(y = 28.94x - 53.38\)

d. What is the slope of the linear regression line? \(\frac{28.94}{1}\)

e. What are the units of the slope expressed as a rate? \(\text{Millions of iPhones sold per year}\)

f. Find the correlation coefficient to the nearest hundredth. Describe the correlation. 0.98; strong positive

g. In January 2016, financial advisors estimated that the global sales for iPhones in 2016 would have dropped to 218.0m. Use your regression line to predict the sales in that year. Is your answer interpolation or extrapolation? Explain. 236.02B; extrapolation; 2016 (year 10) is not in the domain.

h. Do a search to find the actual global sales of iPhones in 2016. How do the financial advisors’ estimate, regression line prediction, and the actual sales compare? Explain. 13 hours is 211.88M
You Write the Story!

Even after home use of the Internet was first introduced in the 1990s, there weren’t a lot of people who had access to the World Wide Web. For many, that access was considered a luxury and most definitely a discretionary expense. As years go by and the Internet becomes a bigger part of our lives, there is a real question as to whether it is no longer discretionary but essential. The graph depicts the estimated number of smart connected devices (Internet-connected devices) in use worldwide by category. Write a short, news-type article centered on the data in this graph. You can find an electronic copy of it at www.cengage.com/financial_alg2e. Copy and paste it into your article.

Examine the equation below used to determine the mean. Look through this chapter and your notes to help you write a problem that could be modeled by the equation.

\[
\frac{4(24.53) + 10(25.00) + 8(27.75) + 3(28.10) + 2(30.00)}{27} = 26.46
\]

1. In Chapter 1 you learned about normal distribution. Do you think the heights of National Basketball Association (NBA) players are normally distributed? In this project, you are going to find out. You will need to get the heights of all NBA players by using the Internet. Then you will need to get your calculator’s instruction manual and look up normal probability plot. You can also search online for a normal probability plot website. A normal probability plot allows you to see if a distribution is normal. Enter the NBA data and interpret the plot and report the mean, median, and standard deviation. Work with a partner since there are more than 400 players. Is the distribution of NBA players approximately normal? Prepare your findings on a poster board including summary statistics and the normal probability plot.

2. A very famous problem in statistics is the “Birthday Problem.” Talk to your math teacher about polling every class in the school that meets during your math period. This will allow you to access to the other classes in the building. Make a list of every class you visit, the number of students in the class, and whether or not that class had in it at least two students with the same
birthdate. What percentage of your classes had a matching birthday? When you have finished, do an Internet search on the solution to the Birthday Problem. Compare it to the data you analyzed. Does the solution defy your mathematical intuition? Explain why or why not.

3. For this project you are going to make a poster board we will call a Balloon Help Tutorial. You are going to pick a problem from Chapter 1, preferably one with several parts to it. You can complete the problem using pen and paper or using a word processor. In either case, leave space on your solution for the text-box “balloons.” Insert text boxes explaining what you did on each step, with arrows pointing to the steps. Use color with discretion, only to help make your explanations clearer. Ask your teacher to make a showcase or bulletin board compiling all of the completed Balloon Help Tutorials.

4. A game show that began in the 1960s titled Let’s Make a Deal inspired a worldwide mathematical controversy and then became a popular math brainteaser. It is called the Monty Hall problem, after its popular host. Do an Internet search on the Monty Hall Problem. Prepare a PowerPoint presentation on the history of the problem and explain its solution. Does the solution defy your mathematical intuition? Explain why or why not.

5. This project is a natural for baseball fans. Do an Internet search to find out exactly what “arbitration” is in baseball labor negotiations. Then pick a current or past player who is/was eligible for arbitration. You are going to use statistics to prepare a case for both sides of the negotiation—the player and the team. Use actual data to see how you can “interpret” the same player’s statistics in ways that aid both parties in the arbitration hearing. Prepare your findings in a PowerPoint presentation.

6. How can you use relative frequency to make predictions? For this problem you will need a plastic or paper cup and a large bag of colored, candy-coated chocolates. Work with a partner. Take two large handfuls of the candies and leave the rest in the bag. Place them in a cup large enough so you can mix up the candies by shaking the cup. You are going to draw a single candy, mark down its color, and put it back and mix them up again. Repeat for a total of 100 draws. Find the relative frequency of greens based on your data. Then, put all the candies back into the original bag with the rest of the candies. Just count the total number of pieces and the total number of greens. Then compute the relative frequency of greens you found from the whole bag. Determine if the relative frequency of greens you found from your repeated 100 draws accurately predicted the relative frequency of greens in the entire bag. Compute the percent error.

7. For this problem you are going to visit your local town hall to secure data on your town. The data can be on all aspects of your town—population, local businesses, schools, income, taxes, recreation, transportation—any aspect that helps describe your town. Prepare your findings on a poster board or as part of a PowerPoint presentation.

8. What is the average weight of a National Football League (NFL) player? Since there are over 1,600 players, you are going to need to take a simple random sample. You will need to get a list of all NFL players by using the Internet. Then you will need to search online for instructions on how to use a random number table. A random number table allows you to number the players and then take a random sample. Take a random sample of 100 players and look up their weights. What is the average weight? How did the random number table help you pick a sample in an unbiased manner? See if you can find, online, the average weight of an NFL player. Compare your sample mean to this population mean and find the percent error.
9. For this project, you are going to get in touch with the school’s parent-teacher organization. They are always looking for information from students and/or teachers and you are going to offer your statistical services to help them. You can help design a survey on a topic of their choice, or use their existing survey and help distribute it and tally the data. You can help them with summary statistics. Interview them to see how you can help, create a plan, and execute that plan. Prepare your findings in a format that is most helpful to the parent-teacher organization.

10. It has been well documented that driving while on a cell phone is distracting, whether or not it is legal in your state. For this project you will need a team of eight or more participants. You are going to pick a busy intersection in your neighborhood. Stand on the sidewalks only! You need one person for each direction cars are moving, to count the number of cars that pass in an hour. You will need a separate person for each direction to count the number of drivers using cell phones as they drive. Spend an hour or more compiling data and create a PowerPoint summarizing all of your statistical findings. Find some Internet research on studies about the dangers of distracted driving and include those findings in your research. Decide as a group if distracted driving is an issue in your community.

11. “What did you do over your summer vacation?” For this project you will create your own survey. Talk to your teacher about a way you can administer it to all students in your school. The survey can be anonymous. Compile the data and prepare your findings on posters for a bulletin board display in your school.

12. Are items sold on Internet “auction sites” always less expensive? This project will help you find out. Search for a product on an auction site by name and/or model number. Be sure you are always comparing the exact same product, size, and quantity. Make a list of the different prices you see for that item in a frequency table. Search for the same product at major stores on their websites and make a frequency table. Compare the auction site prices with the store prices using the two frequency distributions. Include measures of spread since that is the gist of the project.

13. How much do students in your school use their cell phones? Talk to your teacher about polling them to find their data usage or texts for the month. Prepare a handout that tells students where they can find the data you are requesting. Then make a frequency distribution to report your findings. Decide which statistics are important to report. Prepare a poster to show your results.

14. How dispersed are airline prices? Pick a departure and a destination city. Airline prices fluctuate tremendously, depending on the carrier, the trip length, demand for the trip, the time of year, and how far in advance the flight is. Check Internet prices for the same flights by the same carriers each day for 3 weeks. Compile all of your data and report your findings using frequency tables and measures of spread as part of a PowerPoint presentation.

15. In this chapter you learned about planning for personal expenses. Do you have a specific interest related to any of the topics discussed? For this project you will design your own project assignment based on your interests. Read all of the Reality Check projects to get a basic idea of what comprises a project assignment. When you have decided on a project topic and plan, download the project proposal form at www.cengage.com/financial_alg2e and carefully complete the project information. Create a project assignment for a project you would like to complete. You will need to get it approved by your teacher before undertaking the project. Upon teacher approval, complete the project you have created.
Chapter 1
Discretionary Expenses

In the Really? Really! section of this chapter on page 3, you examined how average Americans spend their leisure time. Often, when we look back at our leisure time we view it as “time well spent” because we engaged in activities that made us happy. Let’s look back over the course of an average lifetime to see exactly how much of that well-spent time was used and what was left for other activities.

Extend the Leisure Activity chart to determine how much leisure time is spent over the course of an average life. We will walk you through the steps for filling in the third column of the chart for “watching TV.”

a. According to the chart, the average time spent per day in minutes watching TV is 168.9. Over the course of a year (use 365 days), how much time is spent in minutes watching TV? Express your answer to the nearest tenth. 81,648.5 minutes

b. Convert your answer from part a into hours. Do not round. 1,027.475 hours

c. Convert your answer from part b into days. Round to the nearest tenth. 42.8

This number represents the average number of days per year spent watching TV.

d. The average person lives to 82.5 years of age. How many days of TV are watched in this average lifetime? Round to the nearest day. 3,531

e. Convert the amount you found in part d into years (use 365 days in a year). Round to the nearest tenth. This number represents the average number of years spent over the course of an average lifetime watching TV. 9.7

f. Fill in the remaining amounts in the chart. To save calculating time, steps a through e can be written in one keystroke sequence as follows:

\[
(\text{average minutes spent per day}) \times \frac{365}{24} \times \frac{365}{365} = \text{average minutes per year} \times \frac{365}{24} \times \frac{365}{365}
\]

This will yield the average number of years spent on the activity over the course of 82.5 years. Round all answers to the nearest tenth of a year.

<table>
<thead>
<tr>
<th>Leisure Activity</th>
<th>Average Minutes Spent Per Day</th>
<th>Average Time Spent Over the Course of 82.5 Years (in Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watching TV</td>
<td>168.9</td>
<td>9.7</td>
</tr>
<tr>
<td>Socializing/communicating</td>
<td>42.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Playing games/computer use</td>
<td>26.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Reading</td>
<td>19.6</td>
<td>1.1</td>
</tr>
<tr>
<td>Sports/exercise/recreation</td>
<td>17.9</td>
<td>1</td>
</tr>
<tr>
<td>Relaxing/thinking</td>
<td>16.6</td>
<td>0.95 or one year</td>
</tr>
<tr>
<td>All other leisure activities</td>
<td>24.8</td>
<td>1.4</td>
</tr>
</tbody>
</table>

g. If the average American spends 25 years of a lifetime sleeping and 4 years of a lifetime eating, how much of a lifetime is unaccounted for after you subtract sleeping, eating, and all leisure activities listed in the chart? 35.4 years

Spend your money and your time wisely!
1. Jack wants to buy a pair of name-brand headphones on a popular auction website. He enters the brand, make, and model number and searches the site for what is available. He notices that some headphones are being offered as “buy now” with a fixed price and others are up for bid in auctions. Below is a list of all prices at the time he did his search.

$75  $95  $180  $136  $89  $75  $110  $180  $100  $136  $180  $75  $90

a. Construct a frequency distribution for the data. See margin.
b. Use the frequency distribution to determine the mean. Round your answer to the nearest cent. $123.32
c. Use the frequency distribution to determine the median and the mode. Median: $110; Mode: $180

2. The following table lists the amount of discretionary spending by college students from 2008 to 2013 in billions of dollars.

<table>
<thead>
<tr>
<th></th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>85</td>
<td>89</td>
<td>92</td>
<td>87</td>
<td>121</td>
<td>117</td>
</tr>
</tbody>
</table>

a. Write the formula for the mean yearly college student discretionary spending amount over the 6-year period using sigma notation and determine that mean. Round your answer to the nearest tenth of a billion. $98.5B
b. Write the formula for the mean yearly college student discretionary spending amount for the last 4 years using sigma notation and determine that mean. Round your answer to the nearest tenth of a billion. $104.3B
c. Write the formula for the mean yearly college student discretionary spending for 2009 through 2011 using sigma notation and determine that mean. Round your answer to the nearest tenth of a billion. $89.3B

3. The senior class is planning a graduation trip. The trip committee surveyed the 250 seniors asking the amount they were willing to spend on the entire trip. The survey results are shown in this table.

<table>
<thead>
<tr>
<th>Amount</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0–99.99</td>
<td>15</td>
</tr>
<tr>
<td>$100–199.99</td>
<td>25</td>
</tr>
<tr>
<td>$200–299.99</td>
<td>30</td>
</tr>
<tr>
<td>$300–399.99</td>
<td>72</td>
</tr>
<tr>
<td>$400–499.99</td>
<td>60</td>
</tr>
<tr>
<td>$500–599.99</td>
<td>28</td>
</tr>
<tr>
<td>$600–699.99</td>
<td>20</td>
</tr>
</tbody>
</table>

a. Extend the table to include the columns “Relative Frequency,” “Cumulative Frequency,” and “Relative Cumulative Frequency.” Determine the values under each column heading. Round your answers to three decimal places. See Additional Answers.
b. Using your table from part a, what percent of the senior class were willing to spend between $200 and $299.99? 12%
c. Using your table from part a, how many seniors surveyed indicated that they would be willing to spend up to $499.99? 202
d. Using your table from part a, what percent of the seniors were willing to spend no more than $399.99? 56.8%
e. Since most of the responses were in the $300–399.99 interval, the trip committee made a frequency table of the actual amounts recorded in that interval, as shown here:

<table>
<thead>
<tr>
<th>Amount</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$300</td>
<td>7</td>
</tr>
<tr>
<td>$310</td>
<td>4</td>
</tr>
<tr>
<td>$325</td>
<td>24</td>
</tr>
<tr>
<td>$350</td>
<td>18</td>
</tr>
<tr>
<td>$360</td>
<td>10</td>
</tr>
<tr>
<td>$375</td>
<td>6</td>
</tr>
<tr>
<td>$380</td>
<td>2</td>
</tr>
<tr>
<td>$395</td>
<td>1</td>
</tr>
</tbody>
</table>

What are the percentile ranks for $325 and $350? $325: 48.6%; $350: 73.6%
4. Since vacation travel is a discretionary expense, consumers want to make the most out of the money they have budgeted. They often check travel websites for hotel ratings before making any reservations. A popular travel website offers travelers the opportunity to rate a hotel experience based on five categories: Excellent, Very Good, Average, Poor, and Terrible. Use the spreadsheet to answer the questions below.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rating</td>
<td>Frequency</td>
<td>Relative Frequency</td>
<td>Cumulative Frequency</td>
</tr>
<tr>
<td>2</td>
<td>Excellent</td>
<td>491</td>
<td>0.427</td>
<td>491</td>
</tr>
<tr>
<td>3</td>
<td>Very Good</td>
<td>448</td>
<td>0.390</td>
<td>939</td>
</tr>
<tr>
<td>4</td>
<td>Average</td>
<td>144</td>
<td>0.125</td>
<td>1083</td>
</tr>
<tr>
<td>5</td>
<td>Poor</td>
<td>40</td>
<td>0.035</td>
<td>1123</td>
</tr>
<tr>
<td>6</td>
<td>Terrible</td>
<td>27</td>
<td>0.023</td>
<td>1150</td>
</tr>
<tr>
<td>7</td>
<td>Total</td>
<td>1150</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Write the spreadsheet formula that was used to calculate the amount in cell B7. \[=\text{SUM(B2:B6)}\]
b. Write the spreadsheet formula that was used to calculate the amount in cell C3. \[=\text{B3/B7}\]
c. Write the spreadsheet formula that was used to calculate the amount in cell D5. \[=\text{D4+B5 or SUM(B2:B5)}\]
d. Write the spreadsheet formula that was used to calculate the amount in cell E6. \[=\text{D6/B7}\]
e. Describe what the amount in cell C5 means in the context of this problem.
f. Describe what the amount in cell D4 means in the context of this problem.
g. Describe what the amount in cell E3 means in the context of this problem.

5. The following table shows the number of times Key Club students at Peconic High School went to Echo Beach last summer.

<table>
<thead>
<tr>
<th>Number of Echo Beach Trips</th>
<th>Frequency, f</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
</tr>
</tbody>
</table>

a. How many students are in the Key Club?
b. Find the mean, median, mode, range, variance, and standard deviation for the distribution. Round to the nearest tenth.
c. Millions of visitors go to Echo Beach each summer, many of them repeatedly. If the distribution of the number of visits per person last summer is normal with mean 6.2 and standard deviation 2, find the percent of visitors who went to the beach more than eight times. Round to the nearest percent.
d. Using the statistics from part c, find the percent of visitors who went to the beach between five and seven times, inclusive. Round to the nearest percent.

6. Discretionary music purchases are a big business. Albums can be downloaded, streamed, played from a CD, played on vinyl, and more. What types of albums are being purchased in the United States? Music albums can be classified as current or catalog albums. The difference between the two is that a catalog album is at least 18 months old. Questions 6 and 7 focus on
data comparing album unit sales in the millions in both categories over a period of 7 years, as illustrated in this chart:

<table>
<thead>
<tr>
<th>Year</th>
<th>Year Number</th>
<th>Current Album Sales in Millions</th>
<th>Catalog Album Sales in Millions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>1</td>
<td>209.9</td>
<td>163.9</td>
</tr>
<tr>
<td>2010</td>
<td>2</td>
<td>187.3</td>
<td>138.9</td>
</tr>
<tr>
<td>2011</td>
<td>3</td>
<td>179.4</td>
<td>151.2</td>
</tr>
<tr>
<td>2012</td>
<td>4</td>
<td>161.0</td>
<td>155.0</td>
</tr>
<tr>
<td>2013</td>
<td>5</td>
<td>152.0</td>
<td>136.0</td>
</tr>
<tr>
<td>2014</td>
<td>6</td>
<td>130.5</td>
<td>126.5</td>
</tr>
<tr>
<td>2015</td>
<td>7</td>
<td>118.5</td>
<td>122.8</td>
</tr>
</tbody>
</table>

a. Construct a scatter plot using the year number as the x-coordinate and current album sales as the y-coordinate. Use 0 ≤ x ≤ 8 as the x-axis interval and 100 ≤ y ≤ 220 as the y-axis interval. See margin.

b. Use the scatter plot points to find the equation of the linear regression line for the current album sales. Round the slope and the y-intercept to the nearest hundredth. $y = -14.83x + 221.97$

c. What is the slope of the linear regression line? $-14.83$

d. What are the units of the slope when it is expressed as a rate? in millions per year

e. Based on the linear regression line, how many millions of current albums would you expect in 2016? Round to the nearest tenth of a million. $103.3$ M

f. Is your answer to part d interpolation or extrapolation? Explain. Extrapolation since 8 is outside of the domain

g. Find the correlation coefficient to the nearest hundredth. $r = -0.99$

h. Describe the correlation. Strong negative

7. a. Construct a scatter plot using the year number as the x-coordinate and catalog album sales as the y-coordinate. Use 0 ≤ x ≤ 8 as the x-axis interval and 100 ≤ y ≤ 220 as the y-axis interval. See margin.

b. Use the scatterplot points to find the equation of the linear regression line for the catalog album sales. Round the slope and the y-intercept to the nearest hundredth, $y = -5.83x + 165.37$

c. What is the slope of the linear regression line? $-5.83$

d. What are the units of the slope when it is expressed as a rate? in millions per year

e. Based on the linear regression line, how many millions of catalog albums would you expect in 2016? Round to the nearest tenth of a million. $118.73$

f. Find the correlation coefficient to the nearest hundredth. $-0.83$

g. Describe the correlation. Strong negative

h. Graph the linear regression equations from problems 6 and 7 on the same set of axes. Use 0 ≤ x ≤ 8 as the x-axis interval and 100 ≤ y ≤ 220 as the y-axis interval. What do you notice about the graphs? What do the graphs tell you in the context of the album sales?

8. The table on the right, from the Statistical Abstract of the United States, shows amusement park attendance at the top 15 amusement parks for given years.

<table>
<thead>
<tr>
<th>Year (x)</th>
<th>Amusement Park Attendance at Top 15 Amusement Parks (in thousands) (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>107,348</td>
</tr>
<tr>
<td>2010</td>
<td>109,321</td>
</tr>
<tr>
<td>2011</td>
<td>112,509</td>
</tr>
<tr>
<td>2012</td>
<td>116,420</td>
</tr>
<tr>
<td>2013</td>
<td>119,951</td>
</tr>
</tbody>
</table>

a. Find the mean attendance, in thousands, for the 5 years shown, to the nearest integer. 113,110

b. Find the standard deviation, in thousands, for the years shown. Round to the nearest integer. 4,597

Since the lines intersect, at some point in 2015 the catalog sales appear to be more than the current sales after that point.
c. How many years had attendance within one standard deviation above or below the mean? 3

d. Represent the years using the numbers 9, 10, 11, 12, and 13 to keep computations easier. Look at the scatter plot. Do the data look linear? Yes

e. Find the coefficient of correlation. Round to the nearest thousandth. 0.994

f. Find the linear regression equation. Round to the nearest thousandth. 

\[ y = 3230.5x + 77574.3 \]

g. Use your findings to predict the attendance at the top 15 amusement parks 2 years from the current year.

9. The summer income of the 3,408 students at Van Buren High School last year was normally distributed with mean $1,751 and standard deviation $421.

a. Approximately what percent of the students had incomes between $1,000 and $2,000? Round to the nearest percent. 69%

b. Approximately how many students had incomes less than $800? Round to the nearest integer. 41

10. Splash City Water Park has 14 different rides. Kerry and Kyra are planning a trip to the park. Their tickets are discounted if they agree to only go on eight different rides together. How many different combinations of rides can they choose? 0.75

11. Boaters at Springs Creek Marina have the option to rent fishing rods. The table below shows the number of boat owners that rented fishing rods last Labor Day weekend. The marina wants to compile some summary statistics for the weekend.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Fishing Rod Rentals</td>
<td>Frequency, f</td>
<td>( x_i f )</td>
<td>((x_i - \bar{x})^2 f)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>13</td>
<td>52</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>8 TOTAL</td>
<td>70</td>
<td>140</td>
<td>212</td>
</tr>
</tbody>
</table>

a. How many boat owners were there on that Labor Day weekend? 70

b. Write the spreadsheet formula that was used to calculate the amount in cell B8. =SUM(B2:B7)

c. Write the spreadsheet formula that was used to find the mean. =C8/B8

d. Write the spreadsheet formula that was used to calculate the amount in cell D5. =((A5-C8/B8)^2)*B5

e. Find the variance for the data set. Round to the nearest hundredth. 3.03

12. The number of days each licensed driver in Marion County uses a commuter rail line annually is normally distributed with mean 66 and standard deviation 21.

How many days must a driver use a commuter rail line annually to be in the top 5% of rail users? Round to the nearest integer. 101