Advanced Mathematical Decision Making  
*In Texas, also known as*
Advanced Quantitative Reasoning

Student materials
Semester 1
  Unit I: Analyzing Numerical Data  
  Unit II: Probability  
  Unit III: Statistical Studies

This course is a project of  
The Texas Association of Supervisors of Mathematics and  
The Charles A. Dana Center at The University of Texas at Austin  
*With support from the Greater Texas Foundation*  

2010
Copyright 2010, Charles A. Dana Center at The University of Texas at Austin

This 2010 implementation edition of the AMDM/AQR student materials is copyrighted by the Charles A. Dana Center at The University of Texas at Austin. This February 2011 update includes only minor editorial changes to address the adoption by the Texas State Board of Education of Texas Essential Knowledge and Skills (TEKS) for this course—under the name Advanced Quantitative Reasoning.

Unless otherwise indicated, the materials found in this resource are the copyrighted property of the Charles A. Dana Center at The University of Texas at Austin (the University). No part of this resource shall be reproduced, stored in a retrieval system, or transmitted by any means—electronically, mechanically, or via photocopying, recording, or otherwise, including via methods yet to be invented—without express written permission from the University, except under the following conditions:

1) The following guidelines address only the materials owned by the Charles A. Dana Center at the University of Texas at Austin.
   a) The people of Texas are granted by the University a nonexclusive license in perpetuity—subject to the terms and conditions listed below (item 2)—to use the 2010 AMDM/AQR instructional materials, including these student materials (reflecting improvements based on the 2009-2010 pilot and a minor update based on the adoption of Advanced Quantitative Reasoning TEKS by the Texas State Board of Education), in Texas classrooms and homes for the education of the state’s children.
   b) Teachers and administrators from outside Texas who participate in the Dana Center’s AMDM professional development (registration information available at www.utdanacenter.org/pd) may use the 2010 AMDM/AQR edition of the instructional materials, including these student materials, in their classrooms and schools, subject to the terms and conditions listed below (item 2).
   c) Organizations or individuals other than those listed above (items a and b) must obtain prior written permission from the Charles A. Dana Center at The University of Texas at Austin for the use of the 2010 AMDM/AQR edition of the instructional materials, the terms of which use may be set forth in a copyright license agreement, and which may include the payment of a licensing fee, or royalties, or both.

2) Terms and conditions: Any portion reproduced must retain the original copyright notice, as written above.

We use all funds generated through use of our materials to further our nonprofit educational mission. The Charles A. Dana Center and The University of Texas at Austin, as well as the authors and editors, assume no liability for any loss or damage resulting from the use of this resource. We have made extensive efforts to ensure the accuracy of the information in this resource, to provide proper acknowledgement of original sources, and to otherwise comply with copyright law. If you find an error or you believe we have failed to provide proper acknowledgment, please contact us at dana-txshop@utlists.utexas.edu.

Charles A. Dana Center
The University of Texas at Austin
1616 Guadalupe Street, Suite 3.206
Austin, TX 78701-1222
Fax: 512-232-1855
dana-txshop@utlists.utexas.edu
www.utdanacenter.org
About the Charles A. Dana Center at The University of Texas at Austin

The Dana Center works to raise student achievement in K-16 mathematics and science, especially for historically underserved populations. We do so by providing direct service to school districts and institutions of higher education; to local, state, and national education leaders; and to agencies, nonprofits, and professional organizations concerned with strengthening American mathematics and science education.

The Center was founded in 1991 in the College of Natural Sciences at The University of Texas at Austin. Our original purpose—which continues in our work today—was to increase the diversity of students who successfully pursue careers in science, technology, engineering, and mathematics (STEM) fields.

We carry out our work by supporting high standards and building system capacity; developing collaborations with key state and national organizations to address emerging issues; creating and delivering professional supports for educators and education leaders; and writing and publishing education resources, including student supports.

Our staff of more than 80 researchers and education professionals has worked intensively with dozens of school systems in nearly 20 states and with 90 percent of Texas’s more than 1,000 school districts. As one of the College’s largest research units, the Dana Center works to further the university’s mission of achieving excellence in education, research, and public service. We are committed to ensuring that the accident of where a child attends school does not limit the academic opportunities he or she can pursue.

For more information about the Dana Center and our programs and resources, see our homepage at www.utdanacenter.org. To access our resources (many of them free), please see our products index at www.utdanacenter.org/products. And for updates and background on the Advanced Mathematical Decision Making project, see www.utdanacenter.org/amdm.

About the Texas Association of Supervisors of Mathematics

The mission of the Texas Association of Supervisors of Mathematics is to assist in promoting effectiveness in the supervision, coordination, and teaching of mathematics, especially in the elementary and secondary fields.

TASM accomplishes this by holding meetings for the presentation and discussion of papers; by conducting public discussion groups, forums, panels, lectures, or other similar programs; by conducting or promoting investigations for the purpose of improving the teaching of mathematics; and by the publication of papers, journals, books, and reports, thus vitalizing and coordinating the work of mathematics supervisors across Texas and bringing the interests of mathematics to the attention and consideration of the larger education community in Texas.

For more information about TASM, visit its website at www.tasmonline.net.

About the contents of this course

The materials for the AMDM/AQR course consist of teacher and student materials for Units I through VII. This two-volume resource contains only the student materials. The teacher materials are composed of the student expectations, unit overviews, and unit section planners. Of course, the full 2010 AMDM/AQR instructional materials are available free to the people of Texas, as described in the copyright language above. (The materials are available to educators outside Texas by arrangement; contact dana-txshop@utlists.utexas.edu)
to inquire.) The AMDM website (www.utdanacenter.org/amdm) has information on how Texas educators may obtain the free full set of instructional materials, including the teacher resources.

**About the development of the AMDM course**

The development and production of the original AMDM student expectations, as well as the AMDM instructional materials that constitute Units I through VII of this resource, were supported by the Greater Texas Foundation, based in Bryan, Texas. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Greater Texas Foundation or The University of Texas at Austin.

AMDM student expectations provided the basis for the Texas Essential Knowledge and Skills for Advanced Quantitative Reasoning, which were adopted by the Texas State Board of Education in January 2011. The AQR TEKS have only minor editorial differences from the original AMDM student expectations. Thus, we have updated the AMDM student expectations to match the wording in the AQR TEKS.

- **AMDM student expectations** were developed by a group of 20 people, consisting primarily of Texas mathematics educators and mathematicians at the secondary and postsecondary level, with additional assistance from mathematicians outside Texas. These AMDM student expectations reflect many months of development and feedback from across the state—and from expert reviewers outside the state—as described in updates presented to the Texas State Board of Education in November 2007, June 2008, and March 2009.

- The TEKS for Advanced Quantitative Reasoning have been adopted by the Texas State Board of Education. As such, they are part of state law and are thus available to all Texans at no charge.

- **The AMDM instructional materials** (Units I through VII), now available in Texas as materials for Advanced Quantitative Reasoning, were developed by mathematics teachers and faculty from Texas and beyond, with support from the Texas Association of Supervisors of Mathematics and others in Texas who donated their time or served as consultants for the project. See the acknowledgments section in the teacher materials (available via www.utdanacenter.org/amdm, as described above) for a complete listing of the authors, reviewers, advisory team members, and staff supporting this work.

As noted in the copyright language above, the people of Texas are granted by the University a nonexclusive license in perpetuity to use the 2010 AMDM/AQR instructional materials (reflecting improvements based on the 2009-2010 pilot and a minor update based on the adoption of Advanced Quantitative Reasoning TEKS by the Texas State Board of Education) in Texas classrooms and homes for the education of our children. The full 2010 AMDM/AQR instructional materials are available free to the people of Texas via the AMDM website (www.utdanacenter.org/amdm). The materials are available to educators outside Texas by arrangement; to inquire, contact dana-txshop@utlists.utexas.edu.

This resource was produced in Microsoft Word 2004 for Mac.

Released April 2011.
Advanced Mathematical Decision Making

In Texas, also known as

Advanced Quantitative Reasoning

Unit I: Analyzing Numerical Data

This course is a project of
The Texas Association of Supervisors of Mathematics and
The Charles A. Dana Center at The University of Texas at Austin

With support from the Greater Texas Foundation

2010
Advanced Mathematical Decision Making

*In Texas, also known as*

Advanced Quantitative Reasoning

Student Materials

These student materials are excerpted from one of seven units that make up the 2010 AMDM/AQR instructional materials (developed under the name Advanced Mathematical Decision Making).

**Unit I: Analyzing Numerical Data**

Unit II: Probability
Unit III: Statistical Studies
Unit IV: Using Recursion in Models and Decision Making
Unit V: Using Functions in Models and Decision Making
Unit VI: Decision Making in Finance
Unit VII: Networks and Graphs

Table of Contents

I.A Student Activity Sheet 1: Estimating Crowds .................................................. 1
I.A Student Activity Sheet 2: Filling Your Classroom with Tennis Balls .................. 3
I.A Student Activity Sheet 3: Not Enough Numbers ............................................. 5
I.B Student Activity Sheet 4: Ratios in the Media .............................................. 7
I.B Student Activity Sheet 5: Changing Tires ................................................... 12
I.C Student Activity Sheet 6: Final Grade Averages ........................................... 17
I.C Student Activity Sheet 7: Slugging Averages .............................................. 19
I.C Student Activity Sheet 8: Quarterback Ratings ............................................ 21
I.C Student Activity Sheet 9: Fan Cost Index .................................................... 23
I.C Student Activity Sheet 10: Readability Indices ........................................... 26
I.C Student Activity Sheet 11: Simpson’s Paradox .......................................... 30
I.D Student Activity Sheet 12: Universal Product Codes .................................. 31
I.D Student Activity Sheet 13: Credit Card Numbers ....................................... 34
Analyzing Numerical Data: Estimating Large Numbers
I.A Student Activity Sheet 1: Estimating Crowds

Estimating the number of people in a large crowd (for example, watching a parade or attending/marching in a political rally) is quite challenging and often leads to controversies. One method sometimes used is to focus on a small section of the crowd, such as a rectangular area.

1. Make a square measuring 5 feet by 5 feet, and have your friends stand inside it as if they are watching a band at a small club. Count the number of your friends that comfortably fit in the rectangle and find the ratio of this number to the rectangle’s area. Explain in your own words what this ratio means.

2. Use this value to estimate the size of a crowd that is 10 feet deep on both sides of the street standing along a 1-mile section of a parade route.

3. One rule of thumb for estimating crowds is that each person occupies 2.5 square feet. Use this rule to estimate the size of the crowd watching a parade along the 1-mile section of the route in Question 2.
4. **EXTENSION:** Additionally, a method of estimating the number of people participating in a march is to assign someone to a spot on the sidewalk with the job of counting the number of marchers who pass in 1 minute. Then multiply this number by the minutes it takes for all marchers to pass that spot.

Use this method to estimate the students who attend your school (or a certain building on your campus) on a given morning. Count the number of students entering the building in 1 minute and multiply this number by the minutes it takes the entire group to enter.

5. **REFLECTION:** What other methods could be used to estimate the size of a large crowd? Describe how and why the crowd size estimate supplied by event organizers might differ from estimates made by other groups.
Analyzing Numerical Data: Estimating Large Numbers
I.A Student Activity Sheet 2: Filling Your Classroom with Tennis Balls

1. Consider the situation of determining the number of tennis balls needed to fill your classroom. With a partner or small group, take 5 minutes and make an estimate of the number of tennis balls required to fill the room.

Do you think your estimate is too small or too large? Why?

2. Using the materials provided, make a third estimate. This time, find a lower bound and upper bound for the number of tennis balls you think it would take to fill your classroom. Draw sketches as needed.

3. REFLECTION: In determining your group’s estimate, what mathematical model of a tennis ball did you use? What model of the classroom did you use? Did you make other simplifications or assumptions (for example, assuming there are no desks in the room)?

Fermi Questions

“How large is the crowd?” is an example of a Fermi question, which requires estimating physical quantities to produce an answer. Enrico Fermi was an Italian physicist known for his ability to take limited information and arrive at a reasonable numerical estimate. He approached a problem by setting a lower estimate—a value that the item in question was more than—and an upper estimate—a value that the item in question was smaller than. Fermi then worked through the problem using this approach and reached a numerical estimate within identified limits. Fermi worked at the University of Chicago, and his classic question was “How many piano tuners are there in the city of Chicago?”

Fermi questions rely on making logical assumptions about a situation to reach an estimate to an order of magnitude (usually a power of 10). That is, the estimate is expressed in hundreds, thousands, millions, and so on—whatever unit is appropriate for the question.

Another example of a Fermi question is:

*Can every person in the world fit inside a cubic mile?*

What assumptions do you need to make? What information do you need to find?
Analyzing Numerical Data: Estimating Large Numbers
I.A Student Activity Sheet 2: Filling Your Classroom with Tennis Balls

4. EXTENSION: Tackle one or more of the following Fermi questions:
   a. Which is more valuable: a pile of pennies equaling your weight or a stack of quarters equaling your height?
   b. What is your estimate for the total amount of time students in your school spent during this past semester studying for final exams?
   c. If the land area of Earth was divided equally for each person on the planet, about how much would you get?
   d. How many songs are played on your favorite radio station in a given year?
   e. What is your estimate for the number of airline flights in a year?
   f. What is your estimate for the total number of hairs on your head?
   g. What is your estimate for the number of square inches of pizza consumed by all students in your school during the past year?
   h. How many gallons of gasoline do cars use each year in the United States?
   i. How many hot dogs are eaten at major league baseball games during a season?
   j. How many minutes will U.S. high school students spend texting on their cell phones during the next year?
   k. How much does it cost to raise a child?
Telephone Numbers

During the past few years, there have been several statements of the form: “It is conceivable that we may run out of area codes and telephone numbers within the next 10 years.” The shortage of phone numbers is due to the proliferating use of cell phones, pagers, and fax machines. In the next few problems, you will estimate the number of possible telephone numbers in North America. The assignment of telephone numbers is coordinated by the North American Numbering Plan Administration (www.nanpa.com). (NANPA actually covers only the United States and its territories, Canada, and the Caribbean. Mexico is not included.)

A telephone number in the form NYZ-ABC-XXXX has three sections:

- **NYZ**
  - area code
- **ABC**
  - exchange code
- **XXXX**
  - station code

Before 1995, all area codes had the form NYZ, where N was any digit from 2 to 9 (2-9), Y was 0 or 1, and Z was 1-9 if Y was 0 or Z was 2-9 if Y was 1. The restrictions on N saved 0 for call operator and 1 for long-distance calls. In addition, codes such as 800 and 911 were (and still are) used for special purposes.

The restriction that Y = 0 or 1 was removed in 1995 because all possible area codes had been assigned. Today N is 2-9, Y is 0-8, and Z is 0-9; the exception to these rules are codes of the form 37Z and 96Z, which are being reserved for future use. Area codes where Y = Z are called **easily recognizable codes** and are often assigned to special services such as 800 and 877.

1. How many area codes were possible before 1995?

2. According to the post-1995 rules, how many area codes are possible today?

3. The 7-digit numbers in a given area code have the form **ABC-XXXX**, where X, B, and C can be any digit 0-9 and A is restricted to 2-9. There are two other restrictions:
   - B and C cannot both equal 1 since these values are designated for other purposes such as 911 (emergency) and 411 (information), and
   - 555-0100 through 555-0199 are reserved for fictional uses such as in television shows or movies.

   According to these conditions, how many 7-digit numbers are possible in a single area code?
4. Using your answers to the previous questions, determine how many 10-digit numbers are possible in North America.

5. **EXTENSION:** Some states are running out of license plate numbers. Delaware currently uses six-digit numbers in its license plate numbering system, although recent reports show that its numbers are approaching 1 million (Delaware uses retired numbers for new cars in many cases). The state of Washington recently stated that it needs to explore options to its system of three numerical digits followed by three letters because it is running out of numbers. New Jersey changed its system of three letters followed by two numerical digits and one letter to a new system of one letter, two numerical digits, and three letters. (The last number under the old system was ZZZ 99Z, followed by A10 AAA in the new system.)

   a. How could you determine how many plate numbers New Jersey was able to assign under the previous system? What assumptions did you make in your calculation?

   b. How many additional license plate numbers can New Jersey assign under the new numbering system?

   c. Why do you think the first license plate under the new system was not A00 AAA?

   d. How do New Jersey’s previous and new systems relate to the license plate numbering systems used in Delaware and Washington?

   e. What do license plate numbers tell you about the population in the state?

6. **REFLECTION:** In solving real-world problems such as estimating a crowd size, the number of possible telephone numbers, or the answer to a Fermi question, why is it necessary to first make some reasonable assumptions about the situation?
Analyzing Numerical Data: Using Ratios
1.B Student Activity Sheet 4: Ratios in the Media

For a rectangular shape such as a display screen, the longer side is called the width \((W)\) and the shorter side is the height \((H)\). The aspect ratio is \(W:H\) or \(W/H\).

1. What is the approximate aspect ratio of the screen on your graphing calculator? Consider only the window, not the entire screen.

The size of a television is the length of the diagonal of its screen in inches. The aspect ratio of the screens of older televisions is 4:3, while the aspect ratio of newer wide-screen televisions is 16:9.

2. Find the width and height of an older 25-inch television whose screen has an aspect ratio of 4:3.

Find the area of this screen.

3. Repeat this process to find the width and height of a newer 48-inch television whose screen has an aspect ratio of 16:9.

Determine the area of the screen of a newer 48-inch television whose screen has an aspect ratio of 16:9.
Analyzing Numerical Data: Using Ratios
1.B Student Activity Sheet 4: Ratios in the Media

When movies that were made in one aspect ratio are shown on televisions that have a different aspect ratio, black bars of equal width cover a portion of the screen. Portions of the screen are not needed to project images that were created with different aspect ratios.

Figure 1
4:3 screen displaying a 16:9 image

Figure 2
16:9 screen displaying a 4:3 image

4. Figure 1 shows a letterboxed image with an aspect ratio of 16:9 displayed on a screen with an aspect ratio of 4:3. What percent of the screen’s area is occupied by the image? Justify your answer.

Some people do not like seeing the letterboxes when watching a 16:9 image on a 4:3 display, as shown in Figure 1. What would happen to the image if it filled the height of the TV?

Figure 2 shows a pillarboxed 4:3 image displayed on a 16:9 screen. What percent of the screen’s area is occupied by the image? Justify your answer.

People who own a wide-screen television can choose one of three views of a 4:3 image on their display.

- The normal view shows the pillarboxes, as shown in Figure 2.
- Another option is to stretch the width of the image, keeping the height the same.
- A third option is to zoom in on the image, making the width of the image take the full width of the display.

What affect do these options have on the image?
5. **EXTENSION:** When changing the size of a rectangular image using a computer-drawing program, how can you maintain the same aspect ratio? Investigate this scenario using a drawing program within a computer application.

6. **EXTENSION:** Today most movies made for theaters are shot on film with an aspect ratio of 1.85:1 or 2.35:1. When a film with an aspect ratio of 1.85:1 is shown on a newer wide-screen 16:9 television, minor letterboxing often occurs, but the resulting black bars are usually not visible. Research commonly used aspect ratios and the history associated with this concept.

Why are all aspect ratios identified as something to 1? Or perhaps even recorded as a decimal value?

Express 16:9 to a value in ratio to 1.

A movie with an aspect ratio of 1.85:1 is shown as a letterboxed image on a newer 50-inch 16:9 television. Calculate the height of the image, the height of each barely visible black bar at the top and bottom of the screen, and the percent of the screen’s area that is occupied by the image. Use a variety of representations to justify your response.

A movie with an aspect ratio of 2.35:1 is shown as a letterboxed image on an older 25-inch 4:3 television. Find the height of the image, the height of each black bar at the top and bottom of the screen, and the percent of the screen’s area that is occupied by the image. Use a variety of representations to justify your response.

Based on your analysis of television aspect ratios, why do you think the consumer market has moved having wide-screen TV in the home?
Analyzing Numerical Data: Using Ratios
1.B Student Activity Sheet 4: Ratios in the Media

7. **EXTENSION:** Using appropriate measuring devices, determine the aspect ratio of several objects.

<table>
<thead>
<tr>
<th>Item</th>
<th>Length</th>
<th>Width</th>
<th>Aspect Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheet of paper</td>
<td>11 inches</td>
<td>8.5 inches</td>
<td>11:8.5</td>
</tr>
<tr>
<td>Index card</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A first-class piece of mail (letter or postcard) must have an aspect ratio that falls between 1.3 and 2.5, inclusive. What are some typical sizes of envelopes and postcards? Justify your reasoning.

8. **EXTENSION:** The involvement of aspect ratios in taking and printing photographs is essentially the same as it is in showing movies on a television set. Common still photography aspect ratios are 4:3 (most point-and-shoot digital cameras) and 3:2 (35-mm film). Common U.S. print sizes (in inches) are 4 x 5, 4 x 6, 5 x 7, 8 x 10, and 11 x 14. Verify the following statement regarding image and print size using a variety of representations.

*If your camera uses a 4:3 aspect ratio, but you want a 4 x 6 (3:2) print, many megapixels are wasted (11%).*
Analyzing Numerical Data: Using Ratios
I.B Student Activity Sheet 4: Ratios in the Media

9. **EXTENSION:** Some figures have what is known as a *golden ratio*, which is an aspect ratio that is considered to be visually appealing. Rectangles that have a golden aspect ratio are often called *golden rectangles*. Which of the following rectangles is a golden rectangle?

![Rectangles](image)

The aspect ratio of the second rectangle from the left is the golden ratio. Conduct research on how the value of the golden ratio is found and why figures with this aspect ratio are considered to be visually appealing.

10. **REFLECTION:** How does learning about the various uses of aspect ratio affect your buying habits of products with a visual screen? What types of questions might you ask a sales representative or product help screen online?
Analyzing Numerical Data: Using Ratios
I.B Student Activity Sheet 5: Changing Tires

You have just purchased a new vehicle equipped with factory-installed P245/70R16 tires. You think these tires look too small, so you replace them with P285/75R16 tires. How does this change in tire size affect the accuracy of speedometer and odometer readings? Specifically, your goal is to complete the following statements:

*If your odometer reading is 20000, you have actually traveled ____ miles.*

*If your speedometer reading is 60, your actual speed is ____ miles per hour.*
The calibration of a vehicle’s speedometer and odometer is based on the circumference of the vehicle’s factory-installed tires. For the P245/70R16 tires,

- \( P \) means passenger tire;
- 245 specifies the tire’s width in millimeters;
- 70 is the tire’s aspect ratio—that is, the ratio of the tire’s height to its width reported as a percent; and
- 16 is the diameter of the tire’s rim in inches.
Analyzing Numerical Data: Using Ratios
I.B Student Activity Sheet 5: Changing Tires

1. Fill in the missing information for each tire size. Find the circumference of each tire.

<table>
<thead>
<tr>
<th>Tire</th>
<th>P245/70R16</th>
<th>P285/75R16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width (mm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspect ratio (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (in.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diameter (in.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumference (in.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. After one rotation of the wheel, how many inches further has the truck with the larger tires traveled than the truck with the factory-installed tires?

3. After one rotation of the wheel, the truck with the larger tires has traveled ____ times further than the truck with the factory-installed tires.

4. Use the results from the table in Question 1 to assist in completing the following statements about the truck after the larger tires have been installed on it.

If the odometer reading is 20000, you have actually traveled ____ miles.

If the speedometer reading is 60, your actual speed is ____ miles per hour.

The following principles apply when determining actual distance and speed traveled according to tire size:

Actual mileage = \( k \cdot \) odometer reading (mileage)

Actual speed = \( k \cdot \) speedometer reading (miles per hour)

where \( k = \frac{\text{circumference of bigger tire}}{\text{circumference of factory-installed tire}} \)
5. What is the percent error in the odometer readings? In the speedometer readings?

6. Using the odometer readings in the truck equipped with the larger tires, you determine that the gas mileage is 18 miles per gallon. What is your actual gas mileage in miles per gallon?

7. If you were driving in the truck with the larger tires and the speedometer showed a speed of 65 miles per hour, could you be ticketed for exceeding the 65-mph speed limit by more than 5 mph? More than 10 mph? Justify your answers.

8. REFLECTION: What is the relationship between the ratio of an actual distance to an odometer distance of 1 mile and the ratio of the circumference of a current tire to the circumference of a factory-installed tire?

9. EXTENSION: On your new small car, you replace the factory-installed P185/75R14 tires with slightly larger P205/75R14 tires. Find the missing number in each statement:

   If your odometer reading is 20000 (miles), you have actually traveled ____ miles.

   If your speedometer reading is 60, your actual speed is ____ miles per hour.
10. **EXTENSION**

The ability of birds and airplanes to fly is related to the aspect ratio of their wings.

Since early airplane wings were usually nearly rectangular, the aspect ratio of a wing was the ratio of the span of a wing or airfoil to the chord of a wing, where the span is the maximum cross-stream dimension (longer side) and the chord is the dimension in the stream-wise direction (shorter side), as shown in the figure below. Later, when wings were tapered or had complex platforms, another definition became necessary; currently, the general definition of aspect ratio is $A = \frac{b^2}{S}$, where $A$ is the aspect ratio, $b$ is the span, and $S$ is the area.

![Diagram of aspect ratio](image)

Demonstrate that the current definition of the aspect ratio of a wing coincides with the earlier definition when the wing was rectangular.

Conduct independent research on how the aspect ratio of a wing is related to the flight of airplanes and birds.
When a *weighted average* is applied to a set of numbers, more importance (weight) is placed on some components of the set. Your final average in this class is probably an example of a weighted average.

Consider two grading systems for determining your final class average. Each system is a weighted average of measures that include test grades, final exam grade, homework, and class participation.

<table>
<thead>
<tr>
<th>Grading System I</th>
<th>Grading System II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test average—40%</td>
<td>Test average—60%</td>
</tr>
<tr>
<td>Final exam grade—25%</td>
<td>Final exam grade—15%</td>
</tr>
<tr>
<td>Homework—25%</td>
<td>Homework—15%</td>
</tr>
<tr>
<td>Class participation—10%</td>
<td>Class participation—10%</td>
</tr>
</tbody>
</table>

1. If your values are the following, which grading system do you prefer and why?
   - Test average = 84
   - Final exam grade = 68
   - Homework = 90
   - Class participation = 95

2. If you score 10 points higher on the final exam, how does your final grade average change under each system?

3. If you score 6 points lower on the final exam, how does your final grade average change under each system? Which system is better for you?

4. Use the following information to find your final course average in each grading system:
   - Test grades {80, 74, 82, 88}
   - Final exam grade = 84
   - Homework = 90
   - Class participation = 95
5. Your averages and values are the following:
   - Test average = 85
   - Homework = 90
   - Participation = 95

   What grade do you need on the final exam to earn a final grade average of at least 87 in each grading system?

6. **REFLECTION:** What weights would you assign to each component to set up a grading system? Each weight must be at least 10%. Why do you think your grading system would be fair and effective?
One example of a weighted average in sports is a batter’s slugging average (or percentage) in baseball. The slugging average (SLG) is calculated using the following equation:

\[ SLG = \frac{(1 \cdot S) + (2 \cdot D) + (3 \cdot T) + (4 \cdot HR)}{AB} \]

where \( S \) = singles, \( D \) = doubles, \( T \) = triples, \( HR \) = home runs, and \( AB \) = total at-bats.

Each single has a weight of 1, each double a weight of 2, each triple a weight of 3, and each home run a weight of 4. An at-bat without a hit has a weight of 0.

In his first season with the New York Yankees, Babe Ruth set a record for slugging average that stood for more than 80 years. In 1920, Ruth pounded 172 hits in 458 at-bats. His hits consisted of 73 singles, 36 doubles, 9 triples, and 54 home runs, resulting in a total base count of \((73 \cdot 1) + (36 \cdot 2) + (9 \cdot 3) + (54 \cdot 4) = 388\). When his total number of bases (388) is divided by his total at-bats (458), the result is .847, his slugging percentage for the season. This record was broken in 2001 by Barry Bonds, who had 411 total bases in 476 at-bats for a slugging average of .863. (Statistics from www.baseball-almanac.com)

1. Find the slugging average for a player with the following statistics:
   - \( S = 68 \)
   - \( D = 40 \)
   - \( T = 4 \)
   - \( HR = 16 \)
   - \( AB = 320 \)

2. REFLECTION: Is it possible to have a slugging average of more than 1?

   Theoretically, what is the highest possible value for the slugging average? Could a player ever achieve this value during a baseball season? Give an example or explain why none exist.

3. A slugging average of .500 or higher is considered a sign of an excellent player. Suppose a player had 4 triples and a batting average of .300 in 400 at-bats (batting average = hits/at-bats). Determine a combination of singles, doubles, and home runs that gives this player a slugging average higher than .500.

   Still using 400 at-bats, what is the maximum number of singles this player could have with a slugging average between .500 and .700? Justify your answer.
4. **EXTENSION:** Research the slugging averages of some current-day professional baseball players. Then research the slugging averages of some college players. Based on your findings, which college players should professional scouts be watching? Support your answer with statistics.
The National Football League (NFL) rates quarterbacks for statistical purposes against a fixed performance standard based on the statistical achievements of all qualified pro passers since 1960. This system allows passing performances to be compared from one season to the next.

The following categories are used to compute the quarterback rating:

- percent of completions per attempt (%COMP)
- percent of touchdown passes per attempt (%TD)
- percent of interceptions per attempt (%INT)
- average yards gained per attempt (YD)

(From www.nfl.com/help/quarterbackratingformula)

The following is the formula for compiling the quarterback rating (QR):

\[
QR = \frac{25 + 10(%COMP) + 40(%TD) - 50(%INT) + 50(YD)}{12}
\]

(Note: This formula is subject to a few conditions discussed after Question 2.)

1. For the first two games of the 2008 season, Dallas Cowboys quarterback Tony Romo completed 45 passes in 62 attempts for a total of 632 yards, with 4 touchdowns and 2 interceptions. Verify that Romo’s quarterback rating for those games is approximately 113. Round each value to the nearest tenth.

   Percent of completions per attempt (%COMP) =

   Percent of touchdown passes per attempt (%TD) =

   Percent of interceptions per attempt (%INT) =

   Average yards gained per attempt (YD) =

2. As of 2009, Steve Young has the highest career quarterback rating in NFL history. He completed 2,059 passes in 3,192 attempts for a total of 25,479 yards, with 174 touchdowns and 85 interceptions. Find Young’s career quarterback rating.
Analyzing Numerical Data: Indices Using Weighted Sums and Averages
I.C Student Activity Sheet 8: Quarterback Ratings

The actual formula for the quarterback rating is subject to some conditions, represented below.

\[
QR = \left[ \frac{5(\% \text{COMP} - 30)}{6} \right] + \left[ \frac{10(\% \text{TD})}{3} \right] + \left[ \frac{25(19 - 2(\% \text{INT}))}{12} \right] + \left[ \frac{25(\text{YD} - 3)}{6} \right]
\]

where it is understood that each item in brackets [ ] is truncated to be no smaller than 0 and no larger than 475/12 (giving a maximum rating of 475/3 or 158.3). This rating is rounded to the nearest tenth.

These truncated values mean there is a minimum and maximum value for each component of the rating.

For example, if \( \frac{10(\% \text{TD})}{3} \leq \frac{475}{12} \), then \( \% \text{TD} \leq \frac{475 \cdot 3}{12 \cdot 10} = 11.875\% \).

3. Show that \( 30\% \leq \% \text{COMP} \leq 77.5\% \).

4. If \( c \leq \text{YD} \leq d \), find the values of \( c \) and \( d \).

5. If \( 0 \leq \% \text{INT} \leq e \), find the value of \( e \).

6. Explain why the actual formula and the simplified formula are the same when each component is between its minimum and maximum possible value.

\[
\frac{5(\% \text{COMP} - 30)}{6} + \frac{10(\% \text{TD})}{3} + \frac{25(19 - 2(\% \text{INT}))}{12} + \frac{25(\text{YD} - 3)}{6} = \frac{10(\% \text{COMP}) - 300 + 40(\% \text{TD}) + 25 \cdot 19 - 50(\% \text{INT}) + 50(\text{YD}) - 150}{12}
\]

7. REFLECTION: An index is a numerical scale. Characteristics of an index can be used for the following:
   - to compare variables with one another or a reference number,
   - to give information about general trends, and
   - to help make comparisons and judgments.

   It is often calculated as a weighted sum of various factors resulting in a single summary number. How can the quarterback rating system be thought of as an index?
An index is a numerical scale. Characteristics of an index can be used for the following:

- to compare variables with one another or a reference number,
- to give information about general trends, and
- to help make comparisons and judgments.

It is often calculated as a weighted sum of various factors resulting in a single summary number.

The Fan Cost Index (FCI), compiled by Team Marketing Report, tracks the cost for a family of four to attend a professional sporting event. The FCI includes the prices of 2 average-price adult tickets, 2 average-price child tickets, 4 small soft drinks, 2 small other drinks, 4 regular-size hot dogs, parking for 1 car, 2 game programs, and 2 least expensive, adult-size adjustable caps. The Average Ticket Price in the following tables represents the average cost of a ticket for each member of the family.

The FCI for each Texas team in professional baseball, basketball, and football for 2006 and 2007 is shown in the following two tables. The Soft Drink column includes the price of 1 drink and its size in ounces.

### 2006 Fan Cost Index

<table>
<thead>
<tr>
<th>Team</th>
<th>Avg. Ticket Price</th>
<th>Soft Drink</th>
<th>Other Drink</th>
<th>Hot Dog</th>
<th>Parking</th>
<th>Program</th>
<th>Cap</th>
<th>FCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rangers</td>
<td>15.81</td>
<td>2.75 (16)</td>
<td>6.00</td>
<td>2.50</td>
<td>8.00</td>
<td>5.00</td>
<td>10.00</td>
<td>134.24</td>
</tr>
<tr>
<td>Astros</td>
<td>26.66</td>
<td>4.00 (21)</td>
<td>7.00</td>
<td>4.00</td>
<td>10.00</td>
<td>4.00</td>
<td>11.00</td>
<td></td>
</tr>
<tr>
<td>Spurs</td>
<td>45.88</td>
<td>3.75 (24)</td>
<td>6.25</td>
<td>4.00</td>
<td>10.00</td>
<td>5.00</td>
<td>18.00</td>
<td>283.02</td>
</tr>
<tr>
<td>Mavericks</td>
<td>54.24</td>
<td>2.75 (12)</td>
<td>4.00</td>
<td>4.00</td>
<td>20.00</td>
<td>6.00</td>
<td>15.00</td>
<td></td>
</tr>
<tr>
<td>Rockets</td>
<td>38.64</td>
<td>4.00 (22)</td>
<td>5.75</td>
<td>4.25</td>
<td>15.00</td>
<td>—</td>
<td>20.00</td>
<td></td>
</tr>
<tr>
<td>Cowboys</td>
<td>66.12</td>
<td>3.50 (32)</td>
<td>5.00</td>
<td>3.50</td>
<td>12.00</td>
<td>5.00</td>
<td>10.00</td>
<td>344.48</td>
</tr>
<tr>
<td>Texans</td>
<td>56.97</td>
<td>3.25 (21)</td>
<td>6.00</td>
<td>5.00</td>
<td>15.00</td>
<td>5.00</td>
<td>20.00</td>
<td></td>
</tr>
</tbody>
</table>

(Compiled from www.teammarketing.com)
Analyzing Numerical Data: Indices Using Weighted Sums and Averages
I.C Student Activity Sheet 9: Fan Cost Index

2007 Fan Cost Index

<table>
<thead>
<tr>
<th>Team</th>
<th>Avg. Ticket Price</th>
<th>Soft Drink (size in oz)</th>
<th>Other Drink</th>
<th>Hot Dog</th>
<th>Parking</th>
<th>Program</th>
<th>Cap</th>
<th>FCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rangers</td>
<td>16.47</td>
<td>3.00 (20)</td>
<td>6.50</td>
<td>2.75</td>
<td>8.00</td>
<td>5.00</td>
<td>11.00</td>
<td>141.88</td>
</tr>
<tr>
<td>Astros</td>
<td>26.90</td>
<td>4.00 (21)</td>
<td>7.00</td>
<td>4.25</td>
<td>10.00</td>
<td>4.00</td>
<td>12.00</td>
<td>196.60</td>
</tr>
<tr>
<td>Spurs</td>
<td>51.24</td>
<td>2.50 (24)</td>
<td>6.00</td>
<td>4.00</td>
<td>10.00</td>
<td>6.00</td>
<td>21.00</td>
<td>306.96</td>
</tr>
<tr>
<td>Mavericks</td>
<td>60.58</td>
<td>2.75 (15)</td>
<td>4.00</td>
<td>4.00</td>
<td>20.00</td>
<td>6.00</td>
<td>15.00</td>
<td>339.32</td>
</tr>
<tr>
<td>Rockets</td>
<td>41.98</td>
<td>4.00 (22)</td>
<td>7.00</td>
<td>4.50</td>
<td>15.00</td>
<td>—</td>
<td>20.00</td>
<td>270.92</td>
</tr>
<tr>
<td>Cowboys</td>
<td>84.12</td>
<td>3.50 (32)</td>
<td>5.00</td>
<td>3.50</td>
<td>12.00</td>
<td>5.00</td>
<td>10.00</td>
<td>416.48</td>
</tr>
<tr>
<td>Texans</td>
<td>62.41</td>
<td>3.25 (21)</td>
<td>6.00</td>
<td>5.00</td>
<td>15.00</td>
<td>5.00</td>
<td>20.00</td>
<td>359.64</td>
</tr>
</tbody>
</table>

(Compiled from www.teammarketing.com)

1. Complete the 2006 table by calculating the missing FCI values.

2. There are seven components in the FCI. Find the largest percent increase in a single component from 2006 to 2007 by giving the team, component, and percent increase.

3. Which of the Dallas-area teams (Rangers, Mavericks, Cowboys) had the following during 2006 or 2007:
   a. Most expensive soft drink in terms of price per ounce?
   b. Least expensive soft drink in terms of price per ounce?

4. **EXTENSION:** Create a personal cost index (PCI) in which you select only the items and the number of those items that you and three friends would buy (for example, 4 tickets, 8 hot dogs, no soft drinks, no caps). In 2007, which team is the best buy according to your PCI? Why?
The following table contains the average FCI for all teams in Major League Baseball (MLB), the National Basketball Association (NBA), and the National Football League (NFL) for 2002 through 2007.

<table>
<thead>
<tr>
<th>FCI Average</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLB</td>
<td>145.26</td>
<td>148.66</td>
<td>155.52</td>
<td>164.43</td>
<td>171.19</td>
<td>176.55</td>
</tr>
<tr>
<td>NBA</td>
<td>254.86</td>
<td>261.26</td>
<td>263.44</td>
<td>267.37</td>
<td>274.67</td>
<td>281.90</td>
</tr>
<tr>
<td>NFL</td>
<td>290.41</td>
<td>301.75</td>
<td>321.62</td>
<td>329.91</td>
<td>346.16</td>
<td>367.31</td>
</tr>
</tbody>
</table>

This table contains the rates of inflation from 2003 through 2007.

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflation Rate (%)</td>
<td>2.27</td>
<td>2.68</td>
<td>3.39</td>
<td>3.24</td>
<td>2.85</td>
</tr>
</tbody>
</table>

Use this information and additional calculations to answer the following questions.

5. Did the rate of increase of each sport’s average FCI exceed the rate of inflation for all these years? Find the sport, year, and rate of increase for all cases in which the average FCI rate increase did not exceed the inflation rate.

6. Find the sport and year in which the average rate of increase exceeded the inflation rate by the greatest amount.

7. REFLECTION: Why do you think the FCI is highest for the NFL and lowest for MLB?

8. EXTENSION: There are several indices related to consumer spending. Two of them are the Consumer Price Index (CPI) and the Consumer Confidence Index (CCI). Find out how they are calculated and how they are used.
An index is a numerical scale. Characteristics of an index can be used for the following:

- to compare variables with one another or a reference number,
- to give information about general trends, and
- to help make comparisons and judgments.

It is often calculated as a weighted sum of various factors resulting in a single summary number.

Fog Index

The Fog Index, which was developed by Robert Gunning in 1952, measures the readability of a piece of writing. Calculation of the index produces a number that indicates the years of education a person needs to have to easily understand the text. For example, a passage with a Fog Index of 8 has an eighth-grade reading level. The following table provides more detail on the Fog Index.

<table>
<thead>
<tr>
<th>Use</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writers can use the Fog Index to gauge whether or not their material is suitable for the targeted audience. The ideal Fog Index number for a writing piece for the general population is 7 or 8. Technical material usually has a Fog Index of 10 to 15.</td>
<td>1. Select a block of text that is approximately 100 words.</td>
</tr>
<tr>
<td></td>
<td>2. Compute the average sentence length of the selected passage, dividing the number of words by the number of sentences.</td>
</tr>
<tr>
<td></td>
<td>3. Count words with three or more syllables (that is, complex words). Exclude proper nouns, compound words, words ending in suffixes such as -es, -ed, or -ing that result in a third syllable, terminology familiar to the audience, and hyphenated words. Divide the number of complex words by the total number of words.</td>
</tr>
<tr>
<td></td>
<td>4. Add the average sentence length and the percent of complex words.</td>
</tr>
<tr>
<td></td>
<td>5. Multiply the result by 0.4.</td>
</tr>
</tbody>
</table>

Formula

\[ FI = 0.4 \left( \frac{\text{number of words}}{\text{number of sentences}} + 100 \cdot \frac{\text{number of complex words}}{\text{number of words}} \right) \]

Limitations

Considered an accurate indication of readability, the Fog Index has limitations. For example, some multisyllabic words are not considered to be complex by the general population, but the index does not take this into account.
Analyzing Numerical Data: Indices Using Weighted Sums and Averages
I.C Student Activity Sheet 10: Readability Indices

Flesch-Kincaid Grade Level Index
Another widely used measure is the Flesch/Flesch-Kincaid Readability Tests, which consist of the Flesch Reading Ease (developed by Rudolph Flesch in 1948) and the Flesch-Kincaid Grade Level (developed by John P. Kincaid in 1976). The Flesch-Kincaid Grade Level Index converts the Flesch Reading Ease into a number that indicates the years of education a person needs to have to easily understand the text.

Calculate the Flesch-Kincaid Grade Level Index (\(FKGLI\)) using the following formula:

\[
FKGLI = 0.39 \left( \frac{\text{number of words}}{\text{number of sentences}} \right) + 11.8 \left( \frac{\text{number of syllables}}{\text{number of words}} \right) - 15.59
\]

Declaration of Independence (excerpt)
We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness. — That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed. — That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.

1. Find the Fog Index and the Flesch-Kincaid Grade Level Index of the excerpt of the Declaration of Independence.

Check your answer for the Flesch-Kincaid Grade Level Index using MS Word. (Go to the Help section and look up Readability to find how to access this score.)
Analyzing Numerical Data: Indices Using Weighted Sums and Averages
I.C Student Activity Sheet 10: Readability Indices

One of the most famous speeches in U.S. history is the Gettysburg Address—a 275-word speech that required only a few minutes to deliver.

Gettysburg Address

Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation, so conceived and so dedicated, can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we cannot dedicate—we cannot consecrate—we cannot hallow—this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.

2. Find the Fog Index and the Flesch-Kincaid Grade Level Index for the Gettysburg Address.

   Check your answer for the Flesch-Kincaid Grade Level Index using MS Word.

3. Find the Fog Index and the Flesch-Kincaid Grade Level Index for a passage of around 100 words from one of your favorite books or short stories.

4. Find the Fog Index and the Flesch-Kincaid Grade Level Index for a passage of around 100 words from your science textbook.

5. Write the copy for a radio commercial advertising a real or imagined product. The potential buyers of the product are high school students requiring a Fog Index and a Flesch-Kincaid Grade Level Index of between 10 and 12.
6. **EXTENSION**: There are several indices related to consumer spending. Two of them are the Consumer Price Index (CPI) and the Consumer Confidence Index (CCI). Find out how they are calculated and how they are used.

7. **REFLECTION**: When do you think it would be important to pay attention to the reading level of a passage you write? Think of careers that would be affected by reading levels.
Sabermetrics is the statistical analysis of baseball data. This field of study takes its name from the acronym of the Society for American Baseball Research (SABR). A sabermetrician uses statistics to answer objective questions about baseball players.

1. The Smith brothers (Jo, Bo, and Mo) are talented baseball players. In fact, each one had a .300 batting average going into the final day of the season. During the final game, Jo went 3 for 5, Bo went 4 for 5, and Mo went 5 for 5. Assuming Jo, Bo, and Mo had 240, 420, and 600 official at-bats, respectively, before the final game, who ended the season with the highest batting average? Justify your answer.

2. The previous problem is an example of Simpson’s Paradox—averaging the averages of different populations does not necessarily give the average of the combined population. In this case, each player’s batting average for the entire season does not equal the combination of his batting average for all but the last day and his batting average on the last day. This paradox is named for Edward H. Simpson, who described the phenomenon in his 1951 paper.

Find out more about Simpson’s Paradox and other real-world situations where it occurred. Present your findings to the class.

3. REFLECTION: Have you ever experienced a paradox? Describe the situation.
Identification numbers are present everywhere in society. Today’s identification numbers are more sophisticated than those introduced years earlier (for example, Social Security numbers). Today’s numbers have a check digit to partially ensure that they have been correctly scanned or entered into a computer.

Universal Product Codes (UPCs), typically in the form of barcodes, identify retail products.

The 12-digit UPC barcode consists of three parts:

- manufacturer number,
- product number, and
- check digit.

For example, the manufacturer number for the Dr. Pepper Company is 078000 and appears in the first six digits of all of the company’s product UPC barcodes. GS1, formerly the Uniform Code Council, issues a company this six-digit number. Every item sold by a company requires a different five-digit product number. This includes specific products, their different sizes, their array of colors, their variety of flavors, and other distinguishing features. The last number is the check digit, which guards against entry errors and fraud.

The check digit in a UPC number (that is, the twelfth digit) is determined in the following manner:

- Multiply the first digit by 3.
- Add the second digit.
- Multiply the third digit by 3.
- Add the fourth digit.
- Continue this alternating process for the Digits 5 to 12.
Analyzing Numerical Data: Validating Identification Numbers
I.D Student Activity Sheet 12: Universal Product Codes

The check digit is chosen so that the calculation described previously totals a number whose final digit is 0. In the UPC number \(a_1a_2a_3a_4a_5a_6a_7a_8a_9a_{10}a_{11}d\), the check digit is \(d\), for which the sum

\[3a_1 + a_2 + 3a_3 + a_4 + 3a_5 + a_6 + 3a_7 + a_8 + 3a_9 + a_{10} + 3a_{11} + d\]

ends in 0. In this weighted sum, the weights are: \(\{3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1\}\).

When entering a code number, the single-digit error is most common (for example, keying in 8 instead of 3). Another common error is the transposition error, where the order of two adjacent digits is reversed (for example, writing 83 instead of 38). Systems have been established to detect and correct (when possible) these and other errors almost immediately.

1. Show that 0-58200-48826-5 is a valid UPC number.

2. Show that 0-52200-48826-5 is an invalid UPC number.

   If someone made a single-digit error when entering this invalid number, can you tell which digit is incorrect? Why or why not?

   Change one digit in this invalid number so the resulting number is valid.

3. Determine the check digit \(d\) for the UPC number 38137009213\(d\).

4. Suppose you entered 8 instead of 9 when recording the UPC number 1 55210 02149 6. Explain why the UPC method will detect this error.

   Do you think the UPC method will detect all such single-digit errors? Either give several examples of numbers with an error that will not be detected or explain why you think all such single-digit errors will be detected.

5. Suppose you use the weights \(\{7, 1, 7, 1, 7, 1, 7, 1, 7, 1, 3, 1\}\) instead of \(\{3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1\}\) in the UPC method. Do you think this altered method will detect single-digit errors? Either give several examples of numbers with an error that will not be detected or explain why you think all such single-digit errors will be detected.

6. Suppose you use the weights \(\{2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1\}\) instead of \(\{3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1\}\) in the UPC method. Do you think this altered method will detect single-digit errors? Either give several examples of numbers with an error that will not be detected or explain why you think all such single-digit errors will be detected.
7. When entering UPC numbers, will the error of transposing 2 and 7 be detected? That is, ... 72 ... is recorded as ... 27 .... How about 2 and 6? Make a conjecture about the values of $a$ and $b$ for which the error of transposing $ab$ and $ba$ is detected by the UPC method. Explain your reasoning in making your conjecture.

8. **REFLECTION:** In what ways is the UPC check digit calculation like the activities you explored in Section C, “Indices Using Weighted Sums and Averages”?

9. **EXTENSION**
   - The European Article Numbering Code (EAN 13) is the worldwide standard. Conduct research on how the EAN 13 number is calculated and what types of errors this method detects. Prepare a report and presentation about your findings.
   - The International Standard Book Numbers (ISBNs) used to identify books have a check-digit scheme that detects all single-digit and transposition errors. Conduct research on how the ISBN number is calculated and why this method detects all single-digit and transposition errors. Prepare a report and presentation about your findings.
   - In a vehicle identification number (VIN), the check digit comes in the middle of the number instead of at the end. Conduct research on how the VIN is calculated and what types of errors this method detects. Prepare a report and presentation about your findings.
   - Conduct research on the meaning of various digits in credit card numbers and how check digits are determined. Prepare a report and presentation about your findings.
   - Imagine that you are employed by a company that does not have an identification code number for its employees and products. Prepare a report discussing various numbering alternatives and make a recommendation. Include a check digit in your code number. Indicate how successful your method of determining the check digit will be in detecting single-digit and transposition errors. The company can be your school.
Analyzing Numerical Data: Validating Identification Numbers
I.D Student Activity Sheet 13: Credit Card Numbers

Identification numbers are present everywhere in society. Today’s identification numbers are more sophisticated than those introduced years earlier (for example, Social Security numbers). Today’s numbers have a check digit to partially ensure that they have been correctly scanned or entered into a computer.

Credit cards have 16-digit numbers, of which the first 15 digits identify the credit card and the sixteenth digit is the check digit. The following figure shows the significance of the digits:

1. MasterCard numbers begin with 51, 52, 53, 54, or 55. What is the maximum number of credit cards that MasterCard can issue?
A check digit is used to help validate credit card numbers. The credit card companies use the Codabar method to determine the check digit. This method consists of the following steps:

- Add the digits in the odd-numbered positions and double this total.
- Add the number of odd-position digits that are more than 4 to the total.
- Add the even-position digits.
- Choose a check digit that makes this calculation total a number whose final digit is 0.

Libraries, shipping/receiving companies, and blood banks also use the Codabar method.

2. Show that the check digit \( d \) for the VISA card 4162 0012 3456 789\( d \) is 3.

3. What is the check digit \( d \) for the MasterCard number 5424 9813 2720 008\( d \)?

4. Show that 4128 0012 4389 0110 is an invalid VISA credit card number.

   If someone made a single-digit error when entering this invalid number, can you tell which digit is incorrect? Why or why not?

   Change one digit in this invalid number so the resulting number is valid.

5. The following is another way to explain the Codabar method:

   - If a digit is in an even-numbered position, add it to the total.
   - If a digit is in an odd-numbered position, multiply it by 2. If the product is equal to or greater than 10, subtract 9 from the product. Add this difference to the total.
   - After the first 15 digits have been processed, choose the check digit so that the sum of the 16 digits ends in 0.

   Explain why both methods yield the same check digit.

6. Suppose you entered 8 instead of 9 when recording the credit card number 4001 2560 0196 4310. Explain why the Codabar method will detect this error.

   Do you think the Codabar method will detect all such single-digit errors using a method similar to the one for showing all such errors are detected for UPC numbers? Either write Yes or give several examples of numbers with an error that will not be detected.
7. The Codabar method detects 98% of all transposition errors. There is one pair of numbers, however, that when transposed will not be detected by the Codabar method. Find this pair of numbers and explain why this transposition error will go undetected. (Hint: There are several cases to consider depending on the positions of the transposed digits.)

8. REFLECTION: In what ways is the Codabar method like the activities you explored in Section C, “Indices Using Weighted Sums and Averages”?

9. EXTENSION
   • The European Article Numbering Code (EAN 13) is the worldwide standard. Conduct research on how the EAN 13 number is calculated and what types of errors this method detects. Prepare a report and presentation about your findings.
   • The International Standard Book Numbers (ISBNs) used to identify books have a check-digit scheme that detects all single-digit and transposition errors. Conduct research on how the ISBN number is calculated and why this method detects all single-digit and transposition errors. Prepare a report and presentation about your findings.
   • In a vehicle identification number (VIN), the check digit comes in the middle of the number instead of at the end. Conduct research on how the VIN is calculated and what types of errors this method detects. Prepare a report and presentation about your findings.
   • Retailers use Universal Product Codes (UPC) to identify products. Conduct research on the meaning of various digits in the UPC number and how check digits are determined. Prepare a report and presentation about your findings.
   • Imagine that you are employed by a company that does not have an identification code number for its employees and products. Prepare a report discussing various numbering alternatives and make a recommendation. Include a check digit in your code number. Indicate how successful your method of determining the check digit will be in detecting single-digit and transposition errors. The company can be your school.
Advanced Mathematical Decision Making

In Texas, also known as

Advanced Quantitative Reasoning

Unit II: Probability

This course is a project of
The Texas Association of Supervisors of Mathematics and
The Charles A. Dana Center at The University of Texas at Austin

With support from the Greater Texas Foundation

2010
Advanced Mathematical Decision Making

In Texas, also known as
Advanced Quantitative Reasoning

Student Materials

These student materials are excerpted from one of seven units that make up the 2010 AMDM/AQR instructional materials (developed under the name Advanced Mathematical Decision Making).

Unit I: Analyzing Numerical Data
Unit II: Probability
Unit III: Statistical Studies
Unit IV: Using Recursion in Models and Decision Making
Unit V: Using Functions in Models and Decision Making
Unit VI: Decision Making in Finance
Unit VII: Networks and Graphs

Table of Contents

II.A Student Activity Sheet 1: Using Venn Diagrams .............................................. 1
II.A Student Activity Sheet 2: Using Tree Diagrams ............................................... 3
II.A Student Activity Sheet 3: Using Area Models .................................................. 8
II.A Student Activity Sheet 4: All-American Breakfast Choices ................................ 12
II.B Student Activity Sheet 5: Probability in Games ............................................. 14
II.B Student Activity Sheet 6: Driving and Risk ................................................... 16
II.B Student Activity Sheet 7: Stocks and Risk .................................................... 18
II.B Student Activity Sheet 8: Choosing Classes .................................................. 20
II.C Student Activity Sheet 9: Binomial Probability and Expected Value ..................... 22
II.C Student Activity Sheet 10: Expected Allowance ............................................. 25
Probability: Determining Probabilities
II. A Student Activity Sheet 1: Using Venn Diagrams

Ms. Snow conducted a survey of her homeroom. She asked students what math course and what science course they were taking this semester. Below are the results.

<table>
<thead>
<tr>
<th>Students in Ms. Snow’s Homeroom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra II</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

1. Analyze the data in the Venn diagram and list five facts about Ms. Snow’s homeroom.

2. If a student is selected at random from Ms. Snow’s homeroom, what is the probability that the student is taking Algebra II and Chemistry? Explain your reasoning.

3. If a student is selected at random from Ms. Snow’s homeroom, what is the probability that the student is not taking Algebra II or Chemistry? Explain your reasoning.

4. Find the probability $P($Algebra II or Chemistry$)$. Explain your reasoning.

5. Find the probability of a student taking Chemistry, given that the student is not taking Algebra II, or $P($Chemistry/not taking Algebra II$)$.

Students survey 758 spectators at a national championship tennis match. The survey results indicate the following:

- 421 are male,
- 256 have a two-handed backhand swing, and
- 176 of the people with a two-handed backhand swing are female.

Draw a Venn diagram and label the data.

6. What is the probability that a person selected at random from the survey group is male? Explain your reasoning.

7. What is the probability that a person selected at random from the survey group is female? Explain your reasoning.
Probability: Determining Probabilities
II.A Student Activity Sheet 1: Using Venn Diagrams

8. What is the probability that a person selected randomly from the survey group has a two-handed backhand swing? Explain your reasoning.

9. What is the probability that a person selected randomly from the survey group is a male or has a two-handed backhand swing? Explain your reasoning.

10. What is the probability that a person selected randomly from the survey group does not have a two-handed backhand swing, given that the person is male, or \( P(\text{no two-handed backhand/male}) \)?

11. REFLECTION: Describe the characteristics of a situation that suggest the usefulness of a Venn diagram as a model of the situation.

12. EXTENSION: Describe a situation that could be modeled with a Venn diagram and create the diagram. Use the diagram to determine the probability of at least two events that are possible in the situation.
A church group in Washington state sells pumpkins every year to raise money for the children of their town. This year’s crop, however, produced very small pumpkins. The group decided to construct a corn maze in a field and charge customers to walk through the maze. Customers can only walk forward. If the customers end up at an exit with pumpkins, they win a pumpkin. The church group asked some students to advise it on various possibilities of a customer getting a pumpkin.

Students were shown a simple maze as an example.

1. Make a tree diagram to show the group the possible paths customers might take, entering the maze on the upper, middle, or lower path and proceeding to an exit with or without a pumpkin.

   How is this tree diagram different from others you have worked with before?

   Create-a-Sandwich Menu

<table>
<thead>
<tr>
<th>Bread</th>
<th>Meat</th>
<th>Cheese</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Ham</td>
<td>American</td>
</tr>
<tr>
<td>Wheat</td>
<td>Turkey</td>
<td>Swiss</td>
</tr>
<tr>
<td></td>
<td>Beef</td>
<td>Provolone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Muenster</td>
</tr>
</tbody>
</table>

2. Create a tree diagram showing all possible sandwiches.
Assume that you make all the possible sandwich combinations that you can using one choice from each column of ingredients in the table (bread, meat, cheese). Then someone puts all these different sandwiches in unmarked sacks on the counter. Given this information, answer Questions 3-7?

3. What is the probability you will select a sandwich with white bread? Explain your reasoning.

4. What is the probability you will select a sandwich with American cheese? Explain your reasoning.

5. What is the probability that you will select a sandwich on wheat bread with ham and any cheese? Explain your reasoning.

6. What is the probability you will select a sandwich on white bread that has either beef or turkey and has Provolone cheese? Explain your reasoning.

7. What is the probability you will select a sandwich with neither beef nor Muenster cheese? Explain your reasoning.

8. REFLECTION: Write three probability questions that can be answered using your tree diagram, and then provide the answers.
As president of the high school band, Catrina needs to pick a committee of 2 to accompany her each time she visits middle schools. The director told her that each committee had to consist of 1 boy and 1 girl; 5 boys and 4 girls volunteered to go. To be fair, Catrina makes a spinner with the boys’ names and a spinner with the girls’ names. Each time she schedules a visit, Catrina spins each spinner once to determine who goes with her. If a spinner lands on a line, she spins again.

9. Draw a tree diagram to show all the possible combinations of volunteers who might go with Catrina. How many outcomes are in the sample space?

10. Are all the outcomes equally likely? What would make the outcomes not equally likely?

11. What is the probability that Nathan will be selected? Explain your reasoning. List the possible outcomes for 2-person committees that include Nathan.

12. If Ave decides she cannot go on a visit she is scheduled for, how does this change the probability for other boys or other girls to be selected? Explain your reasoning.
13. **EXTENSION:** Create a scenario for the tree diagram below. Write three probability problems that can be answered using the tree diagram, and then provide the answers.

```
  Start
   /\      1
  /  \  1  2
Red  1  1  2
   /\      1
  /  \  1  1  2
Red  2  1  1
   /\      1
  /  \  1  2  1
Blue 1  1  1
   /\      1
  /  \  1  2  2
Pink 1  1  2
```

14. **EXTENSION:** Mr. Silvas surprises his students with a probability challenge that will determine whether they will take a quiz. He puts three cubes in each of three paper bags: a red cube, a white cube, and a blue cube. He divides the class into four groups of students and tells them that each group will draw one cube from each bag. Each group is challenged to come up with a rule to determine whether they will take the quiz based on if their selection of cubes matches or does not match a criteria they identify in advance. Two groups establish a rule for a cube combination that leads to the outcome of taking the test, and the other two groups establish a rule for a cube combination that leads to the outcome of not taking the test.

a. Draw a tree diagram to show the sample space.
b. Evaluate each group’s decisions. Determine the probability that the outcome chosen by each group will occur.

**Group 1:** If the group ends up with a red cube, a white cube, and a blue cube (order does not matter), its members take the test.

**Group 2:** If the group ends up with at least two red cubes, its members take the test.

**Group 3:** If the second cube selected is white, the group does not take the test.

**Group 4:** If the group ends up with two cubes of the same color, its members do not take the test.

c. Suggest another outcome. What is the probability that your outcome will occur?

d. Which group would you join if you got to choose? Why?
Recall the rules for the pumpkin problem you looked at in Student Activity Sheet 2 with a tree diagram: A customer walks forward through the maze with the possibility of winning a pumpkin; this depends on whether there is a pumpkin at the exit where they come out of the maze. One student, Kyra, draws an area model that demonstrates the probability of getting a pumpkin using this maze.

Kyra explains, “As customers enter the maze, what are the path possibilities? They can take the upper path, middle path, or lower path. These three options lead you to divide the area model into three sections. Next, look at each path and decide how to divide each section. The upper path divides into two paths, the middle path stays one path, and the lower path divides into two paths. Next, decide which part of the model of the maze gets a pumpkin and which part does not.”

<table>
<thead>
<tr>
<th>Path</th>
<th>No pumpkins</th>
<th>Pumpkins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper path</td>
<td>No pumpkins</td>
<td>Pumpkins</td>
</tr>
<tr>
<td>Middle path</td>
<td>Pumpkins</td>
<td></td>
</tr>
<tr>
<td>Lower path</td>
<td>Pumpkins</td>
<td>No pumpkins</td>
</tr>
</tbody>
</table>

The probability of getting a pumpkin is $\frac{2}{3}$, and the probability of not getting a pumpkin is $\frac{1}{3}$. 
1. How does the area model Kyra created compare to the tree diagram from your work in Student Activity Sheet 2?

2. Design another possible maze the group might create, perhaps with more branches, and use an area model to show the possible outcomes. Try out your maze with other classmates to see if they are able to draw an appropriate area model.

Below is a drawing of a second maze the church decided to construct.

3. Use an area model to determine the theoretical probability of a customer taking home a pumpkin.

4. If 50 customers enter the maze, how many pumpkins do you expect to give away? Explain your reasoning.

5. The maze has six exits. If you want to give away a lot of pumpkins, at which three exits do you put the pumpkins? Explain your reasoning. (Hint: Number the exits 1 through 6, and have the area model show where the path ends.)

6. If you do not want to give away too many pumpkins, at which three exits do you put the pumpkins? Explain your reasoning.
7. **REFLECTION:** What would a maze look like with equally likely outcomes? What would the corresponding area model look like? What is an advantage of the area model?

8. **EXTENSION:** Because this year’s maze was such a success, Emma draws a plan for next year. Draw a maze that fits her plan.

- Y—The customer gets a pumpkin.
- N—The customer does not get a pumpkin.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>

a. Find P(Y). Explain your reasoning.

b. Find P(N). Explain your reasoning.
9. **REINFORCEMENT**: You can use an area model to analyze probability situations that involve more than one stage. The following example involves selecting a marble (yellow, red, or blue) from one jar and a cube (yellow, red, or green) from another jar.

```
Possibilities for Jar 1
R  Y  G
R  RR RY RG
B  BR BY BG
Y  YR YY YG
```

```
Possibilities for Jar 2
R  Y  G
```

Each cell represents one outcome.

Compare the following Punnett square to a tree diagram representing the same scenario.

- **a.** Find $P(RY)$.
- **b.** Find $P(\text{at least one red})$.
- **c.** Find $P(\text{both being the same color})$. 
Probability: Determining Probabilities
II.A Student Activity Sheet 4: All-American Breakfast Choices

Fallon’s Bistro

<table>
<thead>
<tr>
<th>Eggs</th>
<th>Meat</th>
<th>Bread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fried eggs</td>
<td>Bacon 0.75</td>
<td>Italian toast</td>
</tr>
<tr>
<td></td>
<td>0.15 Ham</td>
<td>Wheat toast</td>
</tr>
<tr>
<td></td>
<td>0.10 Sausage</td>
<td>White toast</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>Italian toast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wheat toast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>White toast</td>
</tr>
<tr>
<td>Scrambled eggs</td>
<td>0.36 Bacon 0.75</td>
<td>Italian toast</td>
</tr>
<tr>
<td></td>
<td>0.15 Ham</td>
<td>Wheat toast</td>
</tr>
<tr>
<td></td>
<td>0.10 Sausage</td>
<td>White toast</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>Italian toast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wheat toast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>White toast</td>
</tr>
<tr>
<td>Poached eggs</td>
<td>0.15 Ham</td>
<td>Italian toast</td>
</tr>
<tr>
<td></td>
<td>0.10 Sausage</td>
<td>Wheat toast</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>White toast</td>
</tr>
</tbody>
</table>

The American Breakfast
Only $3.99
1. Create a scenario that tells the story of the tree diagram data.

2. Using the tree diagram, answer the following questions.
   a. What do you notice about the percentages attached to the choices in the tree diagram?
   b. If you leave the choices and change their percentages, what does this do to the sample space?
   c. What is true about the sum of the probabilities of each outcome in the sample space?

3. How many combinations of the American Breakfast are possible? Explain your reasoning.

4. What is the probability that the next order of the American Breakfast is fried eggs, bacon, and Italian toast? Where on the tree diagram is this outcome indicated? Explain your reasoning.

5. What is the probability that the next order of the American Breakfast is not scrambled eggs, bacon, and Italian toast? Explain your reasoning.

6. What is the probability that the next customer to get the American Breakfast orders either Italian toast or white toast? Explain your reasoning.

7. REFLECTION: What would an area model look like that models the choices at Fallon’s Bistro?

8. EXTENSION: Consider the two options below. Choose one option, and respond to its questions.
   
   Option 1: If Fallon’s Bistro adds pancakes or waffles as a choice to the American Breakfast, how does this addition affect the sample space? Where did you add the new choice? Why did you add it there? Explain your reasoning.

   Option 2: What happens if hard-boiled eggs are added to the egg choices of the American Breakfast? How does this addition affect the sample space? Explain your reasoning.
Victoria is playing a new video game in which the object is to find hidden treasures. To do so, she must travel through several levels, clashing with guards and watchdogs. In one part of the journey, Victoria must pass through two gates (Gate 1, then Gate 2) to get to the next level.

- The chance that Gate 1 is open is 20%.
- The chance that Gate 2 is open is 30%.
- The game designer has programmed the gates so that the probability of both being open at the same time is 0.1.

Draw a model of the situation to help you answer Questions 1-5. Explain why you chose the particular type of model from among the various probability models.

1. What is the probability that both gates are open when Victoria reaches this part of the game? Explain your reasoning.

2. What is the probability that only Gate 1 is open when Victoria reaches this part of the game? Explain your reasoning.

3. What is the probability that only Gate 2 is open when Victoria reaches this part of the game? Explain your reasoning.

4. What is the probability that neither gate is open when Victoria reaches this part of the game? Explain your reasoning.

5. What is the probability that Victoria finds exactly one gate open?

Victoria encounters another challenge in the game. If she zaps a target in one try, Victoria gets a chance to capture a bonus shield. To capture the bonus shield, she must hit a second target in one try. Victoria can hit a target in one try an average of 60% of the time.

Draw a model of the situations to help you answer Questions 6-8.

6. What is the probability that Victoria hits the first target? Explain your reasoning.

7. What is the probability that Victoria captures the bonus shield? Explain your reasoning.

8. What is the probability that Victoria hits the first target and does not hit the second target to capture the shield? Explain your reasoning.
9. **REFLECTION:** How would you advise your friends who might be interested in playing a new video game?

10. **EXTENSION:** Create two probability situations that use conditional probability. Describe the outcomes for these situations.

11. **EXTENSION:** Create a probability situation that uses compound events. Describe the outcomes for the situation. What type of model do you think will help you answer questions about the situation?
Javier will be a high school senior next year. He wants to get a vehicle to celebrate his graduation. Javier’s mother researched vehicle safety and found that 1 of every 6 teenage drivers was involved in some kind of accident. While talking to his math teacher, Javier mentioned that he did not think the risk was high enough to be concerned. Javier decided to survey 500 students, 230 of whom were male, to help him convince his mother to allow him to get a vehicle. No student has both a car and a motorcycle.

The following are the data from Javier’s survey:

<table>
<thead>
<tr>
<th></th>
<th>Car</th>
<th>Motorcycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males with vehicle</td>
<td>150</td>
<td>23</td>
</tr>
<tr>
<td>Males involved in accident</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>Females with vehicle</td>
<td>225</td>
<td>10</td>
</tr>
<tr>
<td>Females involved in accident</td>
<td>15</td>
<td>4</td>
</tr>
</tbody>
</table>

1. Draw a Venn diagram and a tree diagram of the data.

2. Using the data, what is the probability that Javier will be involved in an accident if he gets a motorcycle? Explain your reasoning.

3. Based on these survey data, Javier told his mother that he only has a 1% chance of getting in an accident. Is he correct? Why or why not?

4. Use your Venn diagram to write three facts that help Javier convince his mother to let him get a vehicle.

5. What probability model would you advise Javier to use when he tries to convince his mother?

6. **REFLECTION**: List some advantages and disadvantages for each type of model used in this problem.

7. **EXTENSION**: Research your favorite vehicle’s safety measures and its likelihood of being involved in an accident. Prepare a short presentation of your findings.
8. **EXTENSION:** Investigate other applications of decision making in situations involving risk—both situations where the risk is known (that is, you have some sort of data to determine mathematically how likely it is to occur) and situations where risk information is not known ahead of time. Examples include purchasing insurance, increasing or decreasing premiums on insurance, and not being eligible for insurance because of high risks. Prepare a short presentation of your findings to share with the class.
Javier will be a senior next year in high school. To celebrate his graduation, his grandmother gave him a sizeable amount of money. Since he has scholarships for college, Javier decided to investigate investing the money in stocks. He talked to a consultant, who explained some of her investment types and their returns as compared to the market average. Javier’s choices were to invest in high- or low-capital stock and domestic or international.

The following is a summary of the data for the last quarter:

<table>
<thead>
<tr>
<th></th>
<th>Domestic</th>
<th>International</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-capital stock</td>
<td>250</td>
<td>40</td>
</tr>
<tr>
<td>Low-capital stock</td>
<td>200</td>
<td>10</td>
</tr>
<tr>
<td>High-capital stock with below-average showing</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>Low-capital stock with below-average showing</td>
<td>15</td>
<td>4</td>
</tr>
</tbody>
</table>

1. Draw a Venn diagram and a tree diagram of the data.

2. Using the data, estimate the probability that a high-cap domestic stock performs at or above market average.

3. What type of stock would you suggest Javier invest in if he wants to limit his risk of the stock performing below market average. Why?

4. Contrast the performance of international and domestic stocks as compared to the market average.

5. Write three facts that help Javier decide in which types of stocks to invest.

   Did you use the Venn diagram or the tree diagram to find the facts? Which was most helpful? Why?

6. **REFLECTION**: List the advantages and disadvantages of the types of models used in this problem.
Probability: Everyday Decisions Based on Probabilities
Student Activity Sheet 7: Stocks and Risk

7. **EXTENSION:** Javier decides to invest $100 in high-cap domestic stocks. He narrows his choice to stock in Ttxtel or stock in TxMart, both of which have the same price today. Ttxtel is a growth stock in the technology sector. It tends to be more volatile. Based on financial analysis, this stock may go down $10 with a probability of 0.4 or up $8 with a probability of 0.6 in the next quarter. TxMart is a value stock in the domestic goods sector. It is projected to drop $4 with a probability of 0.2 or go up $2 with a probability of 0.8. Which stock do you recommend? Why?

8. **EXTENSION:** Investigate other applications of decision making in situations involving risk—both situations where the risk is known (that is, you have some sort of data to determine mathematically how likely it is to occur) and situations where risk information is not known ahead of time. Examples include purchasing insurance, increasing or decreasing premiums on insurance, and not being eligible for insurance because of high risks. Prepare a short presentation of your findings to share with the class.
Probability: Everyday Decisions Based on Probability
II.B. Student Activity Sheet 8: Choosing Classes

Four teachers offer Zane’s favorite computer class at different times during the day. The school counselor asks Zane if he prefers a morning or afternoon class. Below is a list of teachers and the periods they teach this class. The morning classes are 1st, 2nd, and 3rd periods, and the afternoon classes are 4th, 5th, and 6th periods.

- Mr. Nelson—2nd, 4th, 5th, 6th
- Ms. Trevino—1st, 2nd, 3rd, 4th, 5th
- Mr. Garza—1st, 3rd, 4th, 6th
- Ms. Jones—1st, 2nd, 3rd, 6th

1. Before answering the counselor’s question, Zane wants to list all the possibilities so he can make a choice that gives him the highest probability of getting a teacher he prefers. Create a table and a tree diagram that illustrate the possibilities.

2. Before deciding on a morning or afternoon class, Zane remembers he wants to take his math class during 3rd period. What is the probability that he will be assigned the computer class during this time?

3. Zane prefers to be in the class of Ms. Trevino or Mr. Nelson. Should he pick the morning or the afternoon? Explain your reasoning.

4. After checking the schedule, the counselor told Zane that Mr. Garza’s classes are filled. How does this information affect the probability of Zane getting any afternoon class? If Zane asks for an afternoon class, how does this affect his probability of getting Mr. Nelson or Ms. Trevino?

5. Because this is a required class for all students and Mr. Garza’s classes are filled, the school adds another teacher, Ms. Lopez. She will teach 1st and 6th periods. Does this fact affect the probability of getting Mr. Nelson in the morning?

6. While Zane is calculating probabilities so he can make his decision, the class offerings change. (Mr. Garza’s classes fill, and Ms. Lopez is added.) If Zane requests an afternoon class, what is his probability of getting Ms. Trevino for 4th period?

7. REFLECTION: What other factors might influence Zane’s class selection? Would these factors be reflected in the diagram or data? If so, describe the new diagram or set of data.
8. **EXTENSION:** Meet with the person on your campus who is responsible for creating the master schedule. Prepare a set of questions similar to the ones that Zane faced and record the responses. Share your findings with the class, including at least one decision you can make based on those findings.
At the National Baseball Batting Contest, the organizers have set up game booths for the contestants. Richard wants to win a large stuffed animal. The rules of the game are as follows:

- You are pitched 5 fastballs, and you must hit them into a fair zone to count.
- If you successfully hit all 5 pitches, you win a large stuffed animal.
- If you successfully hit 3 or 4 pitches, you win a small stuffed animal.
- If you successfully hit 1 or 2 pitches, you win a bat-shaped pencil.
- If you miss all the pitches, you do not win a prize.
- The game costs $3 to play (each set of 5 fastballs).

1. What are the possible outcomes that Richard can have on his 5 swings? Explain or show how you arrived at your answer?

2. What is the probability of Richard successfully hitting the following? Justify your answers.
   a. All 5 pitches?
   b. 4 pitches?
   c. 3 pitches?
   d. 2 pitches?
   e. 1 pitch?
   f. 0 pitches?

3. Suppose Richard tries the game 10 times. How many times do you expect him to successfully hit the given amounts of times below? Justify your answers.
   a. 5
   b. 4
   c. 3
   d. 2
   e. 1
   f. 0
4. Everyone has the same chance as Richard to win. If 160 people each play the game once, how many large stuffed animals will be won? Explain your reasoning.

5. If 160 people each play the game once, how many small stuffed animals will be won? Explain your reasoning.

6. If 160 people each play the game once, how many bat-shaped pencils will be won? Explain your reasoning.

7. If a large stuffed animal costs the game organizers $6, a small stuffed animal costs $1, and a pencil costs $0.25, how much profit do you expect the group to make from the 160 players?

8. EXTENSION: For the game described in which you are pitched 5 fastballs, you found the probabilities of successfully hitting 5, 4, 3, 2, 1, or 0 pitches. Suppose the game is changed and you are pitched various number of pitches. Find the probabilities of each outcome and look for a pattern of numbers. (Because you are looking for a pattern, do not simplify the fractions.)

   1 pitch:
   2 pitches:
   3 pitches:
   4 pitches:

   Consider the geometric arrangements known as Pascal’s triangle below. Decide how the values in the triangle are generated. Then compare the values you found for the probabilities above to the values in the triangle. What do you notice?

   
   1
   1 1
   1 2 1
   1 3 3 1
   1 4 6 4 1
   1 5 10 10 5 1

   If you had to successfully hit 6 of 6 pitches to win the large stuffed animal, what is the probability of winning the large stuffed animal? Use Pascal’s triangle in your solution.
9. **EXTENSION:** Research the use of Pascal’s triangle. Prepare a short class presentation regarding your findings. Be sure to include the connection to combinatorics.

10. **EXTENSION:** Investigate the probabilities and associated expected values related to winning a state or local lottery game. Prepare a short class presentation about the probable outcomes of event of winning the lottery.
Yvonne gets $15 a week for allowance. She also loves to play basketball. Yvonne wants to convince her father to try something new with her allowance based on her basket-shooting talent, hoping that it will increase what she receives. Yvonne suggests that instead of getting $15, she attempt shooting baskets each week for her allowance.

- If she misses the first basket, she gets only $5.
- If she makes (succeeds with) the first basket, she gets $15 and a chance to make another basket for an additional $10.

Yvonne can make a basket 40% of the time

1. To help Yvonne’s father decide whether to use the new allowance plan, find the probability of Yvonne making 0 baskets, 1 basket, and 2 baskets. Justify your reasoning with an appropriate model.

2. How many times in a year do you expect Yvonne to get $5? $15? $25?

3. How much allowance should Yvonne expect to receive in a year?

4. Should Yvonne’s father accept the deal? Justify your reasoning.

Yvonne practiced shooting baskets all year. She can now make a basket 60% of the time. Yvonne offers her father the same deal.

5. What is the probability of Yvonne making 0 baskets? 1 basket? 2 baskets?

6. How much money do you expect Yvonne to receive in a year?

Reinforcement

Yvonne’s younger sister, Lisa, wants her father to offer her the same deal. Lisa can make a basket 20% of the time.

8. You have now looked at Yvonne’s situation (with a 40% shooting percentage) and Lisa’s (with a 20% shooting percentage) to determine what percent of the time they are each likely to receive $5, $15, and $25 in allowance. Make a table and a graph that show the relationship between these shooting percentages and the amount of allowance earned, as well as other possible shooting percentages related to the amount of allowance earned. Describe how this relationship changes based on a person’s shooting percentage.

<table>
<thead>
<tr>
<th>Percentage of Making a Basket</th>
<th>Amount of Allowance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$5</td>
<td>$15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
9. Lisa’s ability to make a basket improved to 30%. Using your graph, what average weekly allowance should she expect? What is the actual average amount Lisa should expect?

10. If Yvonne wants to earn an average weekly allowance of $15, what percentage of baskets must she make?

11. Yvonne’s father figured out that he had been giving Yvonne $20 per week for her allowance. What percentage of the baskets is Yvonne making? What information did you use to answer this question?

12. EXTENSION: Research popular basketball players or teams to find current percentages associated with making free throws, 2-point shots, and 3-point shots. Create a geometric probability model of a scenario using the player or team to make the given shot in a particular situation. Provide explanations as needed for sharing with others in the class.

(Note: Basketball statistics do not break out the 2-point percentage in its own category in the same way that free throws and 3-pointers are broken out. Instead 2- and 3-point goals are combined in the Field Goals Made [FGM] category. Determine the number of 2-point shots made by subtracting the 3-point shots made from the FGM.)

13. MINI-PROJECT: Design a carnival game that is of interest to your age group.
   - Make a scale model of the game.
   - Include a report with your model that details the following:
     a. Experimental and theoretical probability of winning the game
     b. Expenses
     c. Expected payoff
     d. Profit
     e. Rules of the game
     f. Prizes and how to earn the prizes
     g. Why the game should be selected for the carnival
Advanced Mathematical Decision Making

In Texas, also known as

Advanced Quantitative Reasoning

Unit III: Statistical Studies

This course is a project of
The Texas Association of Supervisors of Mathematics and
The Charles A. Dana Center at The University of Texas at Austin

With support from the Greater Texas Foundation
Advanced Mathematical Decision Making

In Texas, also known as
Advanced Quantitative Reasoning

Student Materials

These student materials are excerpted from one of seven units that make up the 2010 AMDM/AQR instructional materials (developed under the name Advanced Mathematical Decision Making).

Unit I: Analyzing Numerical Data
Unit II: Probability
Unit III: Statistical Studies
Unit IV: Using Recursion in Models and Decision Making
Unit V: Using Functions in Models and Decision Making
Unit VI: Decision Making in Finance
Unit VII: Networks and Graphs

Table of Contents

III.A Student Activity Sheet 1: Overview of Purpose, Design, and Studies ............... 1
III.A Student Activity Sheet 2: Treatment of Subjects ........................................... 11
III.A Student Activity Sheet 3: Margin of Error ..................................................... 14
III.A Student Activity Sheet 4: Sampling Design and Methods ............................. 17
III.B Student Activity Sheet 5: Histograms .......................................................... 25
III.B Student Activity Sheet 6: Analyzing Graphical Displays .............................. 31
III.B Student Activity Sheet 7: Using Technology ................................................ 37
III.B Student Activity Sheet 8: Survey Design ..................................................... 45
III.C Student Activity Sheet 9: Introduction to Statistical Bias and Variability .......... 47
III.C Student Activity Sheet 10: Statistical Bias in Research Studies and Polls ........... 51
Music is a large part of many people’s lives. Because of this, it is often the subject of study. For example,

- Music industry executives want to know what will be popular with different age groups.
- Advertisers want to know which radio stations are the most popular.
- Doctors want to know how much hearing damage results from loud music.
- Teachers want to know whether or not listening to classical music helps students perform better on tests.

Statistical investigations are used every day for a variety of reasons.

1. What are the purposes of statistical investigations? Give some examples of statistical investigations with which you are familiar.

This graphic illustrates the process of planning and implementing a statistical investigation. First, a question (or a series of questions) sparks the interest of a researcher. The research team then decides on the best design for investigating the question.

2. The graph shows no obvious ending point (or starting point). What does this mean?
Consider the following examples of two different types of statistical investigations.

Example 1

Radio rating services sometimes collect data on listenership by asking participants to record the date, time, and station each time they listen to the radio. Other rating services distribute monitoring devices that automatically record this information anytime the participant has the radio turned on. Still others call participants and ask them about their listening habits. The data are then compiled so that advertisers know which stations are the most popular at specific times during the day.

Each of these approaches is an example of an observational study, which collects data about some characteristic(s) of the population. The data can be collected by observation, by a survey or interview, or by other means.

3. Describe an observational study in your own words:

   An observational study is research in which

Example 2

A 17-year-old student designed a science fair project with 72 mice randomly assigned to three groups: hard rock music, Mozart, and no music at all (called a control group). The mice in the first two groups were exposed to music 10 hours a day. Three times a week, all of the groups were timed as they ran through a maze. An analysis of results showed that the 24 mice in the no-music group averaged about a 5-minute improvement in their maze completion time, while the Mozart mice improved 8.5 minutes. The hard rock mice actually got slower—an average of four times slower! Another interesting fact: The student had to start his experiment over because all the hard-rock mice killed each other. None of the classical mice did that. (Wertz, M. [1998]. Why classical music is key to education. from www.schillerinstitute.org/programs/program_symp_2_7_98_tchor_.html#Music_Mice_Mazes)

This is an example of an experimental study. In an experimental study, the researcher separates the participants into one or more groups and applies some sort of treatment. After treatment, the variable of interest is measured and the results are compared.

4. What are the treatment and the variable of interest in this case?

5. Describe an experimental study in your own words:

   An experimental study is research in which
Observational and experimental studies have many components that must be planned, such as sampling and data collection procedures. Then the data must be collected, the results analyzed, and the conclusions reported.

6. Referring back to the Research Cycle graphic, why is there an arrow after the Report box?

And what about the Question box? Consider this situation: “This unopened bag of chips is half empty. I wonder if it really contains 28.3 grams as the package says?”

This type of informal question or observation is the beginning of many investigations. Informal questions can turn into more formal problem statements or research questions. For example, you may decide to investigate whether there is a scandal in the potato chip industry by checking the following:

“Do Spud Potato Chips contain an average of 28.3 grams of chips per bag?”

7. REFLECTION: Now that you have been introduced to the research cycle process, think of some research questions that you are interested in studying. List at least three ideas of research questions. Consider the following:

- What type of study (experimental or observational) might be best to approach each of your research questions?
- If you only have four weeks to actually implement a research study, is it still possible to study any of your research questions?
- How can you change one of your questions to make it fit into this timeline?

8. Suppose you conduct the investigation into Spud Potato Chips and find that the mean weight of the chips in your sample is 25 grams, rather than 28.3 grams (\( \bar{X} = 25 \) grams). Do you think that a difference of 3.3 grams between the actual and advertised weights is large enough that it needs to be reported? If so, how do you report this information and to whom?
In some situations, researchers are even more formal and state **hypotheses**. In a case like this, the **null hypothesis** \( (H_0) \) generally states that there is no difference between the true value and the claimed value. The **alternative hypothesis** \( (H_a) \) states that something is different or incorrect, or that something has changed.

9. What are the null and alternative hypotheses for the potato chip example?

   - \( H_0 \): The true mean weight ____________________________
   - \( H_a \): The true mean weight ____________________________

Notice that the hypotheses say “The true mean weight.” This implies that the statements refer to the **population** of all Spud Potato Chip bags, not just a single bag or even a small sample. When a statistical investigation is conducted, it generally employs a sample that is then used to make a generalization about the population. Notice that in this case (as in many cases), population does not refer to people, but to bags of potato chips.

To be concise, researchers often use symbols in place of words. Greek letters are usually used when referring to populations (the entire group being studied, from which a sample or samples will be drawn). English letters are used for samples (the particular items or individuals included in a particular study). For example, when discussing the mean:

   - \( \mu \) = the population mean (Greek letter \( mu \)—pronounced mew)
   - \( \bar{x} \) = the sample mean (pronounced \( x-bar \))

So the hypotheses for a study can be stated in words or symbols. When using symbols, you must identify what your symbols represent.

   - \( H_0 \): \( \mu \geq 28.3 \) grams, where \( \mu \) is the true mean weight of a bag of Spud Potato Chips
   - \( H_a \): \( \mu < 28.3 \) grams

Statistical studies are designed with carefully selected measures that ensure (within error margins) that, if the sample is well selected and the study is well designed and conducted, the mean and other measures of the sample are likely to be similar to the corresponding measures of the population being studied. Sometimes, if the population is small (such as high school seniors in a small town), it may be possible that the sample studied is the entire population. However, often a sample is a smaller subset of a population (such as a research question that might target the entire population of high school seniors in a state or in the nation).
For Questions 10 and 11, practice writing hypotheses. Write them in words and then convert them to symbols. Finally, sketch or outline a simple study design that might help study the hypotheses.

10. A local pizza shop advertises “an average delivery time of 20 minutes or less,” but it does not offer a guarantee such as a free pizza. The national manager, Su Lin, wonders if her employees are fulfilling the claim.

11. James believes that his mother’s houseplants would grow taller if she watered with rainwater instead of tap water.

12. REFLECTION: Recall the potato chip hypotheses:
   - \( H_0: \mu \geq 28.3 \text{ grams} \), where \( \mu \) is the true mean weight of a bag of Spud Potato Chips
   - \( H_a: \mu < 28.3 \text{ grams} \)

   What would you do next to determine which of these hypotheses is true?
The following cases are examples of observational studies and experimental studies. Consider the type and design of each study.

13. Identify the type of study for each case.

- How do you know?
- What is the variable of interest in each case?
- What are some advantages and disadvantages of each teacher’s plan?

a. Mrs. Johnson teaches American History and wanted to help her students do their best on exams. After failing to find any research on different test formats, she decided to conduct her own research. She flipped a coin for each student in her classes. If the coin landed heads up, the student took a multiple-choice test. If the coin landed tails up, the student received a fill-in-the-blank exam. Afterward, Mrs. Johnson compared the averages for the two test formats.

b. In World History, Mr. McDonald had a similar concern. He decided, however, to ask his students. He put a question at the bottom of an exam: “Which do you prefer, multiple-choice or fill-in-the-blank questions?” Afterward, Mr. McDonald tallied the results.

c. Mr. Mitchell was interested in the effects of music on student performance. At the bottom of his exam, he asked students to circle their favorite type of music: rock, country, or hip-hop. He then computed the averages for the students who liked each type of music and compared the results.

d. Mrs. Knox’s senior English classes were working on their term themes. During 2nd period, she allowed students to listen to their choice of music through earphones while working, but her 4th-period class was required to work quietly without music. Mrs. Knox averaged this major grade for each class and compared 2nd period’s average to 4th period’s average.

e. Mr. Paul, the guitar teacher, sat at the food court in the mall and made a tally sheet that noted each t-shirt he saw with a musical group illustrated on it. He compiled the results and posted an entry to his blog about the most popular groups.
14. In the experimental studies:
   - Describe the treatment(s)
   - Who were the participants?
   - How was the assignment of treatment(s) accomplished?

15. A group of participants that the treatment group is being compared to is called the **control group**. Give an example of a treatment, the treatment group, and the control group.

Researchers are often concerned that participants in a study show improvement simply because they are in the study and not because they are receiving an effective treatment. This is called the **placebo effect**.

**Example 1:** Half of the participants in a study for a new headache remedy receive the new pill, while the remaining participants receive a pill containing only inactive ingredients. Participants receiving the inactive pill (the placebo) report that their headaches have been somewhat relieved. These participants believed they were being treated, and this belief may have affected their perception.

**Example 2:** Half of the athletes in a study received a new lotion for strained muscles, while the other half received a lotion with only inactive ingredients. Both groups report improvement in their muscle pain.

16. The improvement in Example 2 could be psychological, as in Example 1. Can you think of any other reason for the improvement?

17. **REFLECTION:** Consider Mrs. Johnson and Mr. McDonald’s exam situations. Suppose Mrs. Johnson’s results overwhelmingly favor fill-in-the-blank exams and Mr. McDonald’s results strongly favor the multiple-choice format. Are these results in conflict with each other? What could be the cause(s) of this difference?

Suppose Mrs. Johnson takes her results to the school board and asks the board to require that all teachers in the district give fill-in-the-blank exams. If a school board member asks you what should be done, what recommendations would you give?

18. **REFLECTION:** Mrs. Johnson applied two different treatments to the participants in her study—some students received a multiple-choice test, and the rest received a fill-in-the-blank exam. Can you think of two treatments that could be used in a medical experiment? In a cooking experiment? Would a situation with three or more treatments be possible? Explain your thinking.
19. **EXTENSION:** Read the summary of actual studies on the following pages. For one of the studies, determine the following information. Be prepared to share your findings in a short presentation.
   a. Is the study observational or experimental? Explain your answer.
   b. Who/what are the participants or experimental units?
   c. If experimental, what was the treatment(s)?
   d. If there was a treatment, how was it assigned?
   e. Was there a control group and/or a placebo?
   f. If observational, what was being observed and why?
   g. Are there any statements that appear to be opinions?
   h. Are there any stated or implied limitations to the study?
Tobacco and middle school and high school students

The Centers for Disease Control and Prevention conducted the National Youth Tobacco Survey (NYTS) in 2004 to measure current use of tobacco products and selected indicators related to tobacco use, including youth exposure to tobacco-related media and access to cigarettes.

The survey was distributed to 267 U.S. public and private schools; 14,034 middle school students and 13,738 high school students completed the survey. Participation was voluntary and anonymous, and school parental permission procedures were followed. Some results included:

A. 11.7% of middle school students and 28% of high school students reported current use of a tobacco product.

B. 77.9% of middle school students and 86.5% of high school students reported seeing actors using tobacco on television or in movies.

C. 70.6% of current cigarette smokers in middle school and 63.9% in high school said they were not asked to show proof of age when they purchased or attempted to purchase cigarettes from a store.

These results indicate very little change from the results of the 2002 survey. The lack of substantial decreases in the use of tobacco products among students indicates the need to:

- increase the retail price of tobacco products,
- implement smoking-prevention media campaigns, and
- decrease minors’ access.

The findings in this report are subject to limitations. First, these data apply only to youths who attended middle school or high school. Among 16- and 17-year-olds in the United States, approximately 5% were not enrolled in a high school program and had not completed high school in 2000. Second, the questionnaire was offered only in English. Thus, comprehension might have been limited for students with English as a second language.

(Adapted from Centers for Disease Control. Tobacco Use, Access, and Exposure to Tobacco in Media Among Middle and High School Students—United States, 2004. from www.cdc.gov/mmwr/preview/mmwrhtml/mm5412a1.htm)
Scientists in Turkey undertook a study of people with epilepsy. The research included analyzing scalp hair samples from 22 participants with epilepsy and 23 participants without epilepsy, checking for differences in levels of copper, iron, zinc, magnesium, and calcium. (The researchers speculate that such differences could indicate metabolic differences that may contribute to epilepsy.) Results indicated that the epileptic group had significantly lower levels of copper and iron compared to the nonepileptic group.


Scientists in Nigeria, in an effort to find an inexpensive method of raising rabbits for food, designed a study to test the effect of replacing some of the rabbits' soybean diet with Gliricidia sepium Leaf Meal (GLM). Twenty-five young rabbits were randomly assigned to receive either 0%, 5%, 10%, 15%, or 20% GLM. The groups showed no significant difference in the amount of harvestable meat, while decreasing costs of raising the meat.

There are two ways to categorize data according to its source—as primary data or secondary data.

**Primary data** are data that you collect directly. Sources of this type of data include the following:

- experiments,
- survey instruments, and
- observation instruments.

When collecting primary data, most researchers run a **pilot study** first. This is a small-scale version of the research plan. What are some reasons to run a pilot study?

**Secondary data** are data that have been collected by someone else and are available to the researcher. The following are examples of this type of data:

- the Internet, including websites such as the U.S. Census Bureau and other government sites;
- printed materials, such as books, almanacs, newspapers, and magazines; and
- historical documents.

When using secondary data, part of the research plan must include ensuring that the data are reliable. When collecting primary data, you are responsible for doing everything possible to ensure that your participants are well informed and safe.
One nice thing about a potato chip study is that researchers do not have to worry about hurting the chips, and the chips’ permission is not needed! With human participants, you have to be more careful.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was created by the 1974 enactment of the National Research Act. Its purpose is “to identify the basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects and to develop guidelines which should be followed to assure that such research is conducted in accordance with these principles.” (from Office of Human Subjects Research on the National Institutes of Health website: ohsr.od.nih.gov/guidelines/belmont.html)

The commission developed the following guidelines:

A. Ethical Principles and Guidelines for Research Involving Human Subjects
B. Boundaries Between Practice and Research
C. Basic Ethical Principles
   1. Respect for Persons
   2. Beneficence
   3. Justice
D. Applications
   1. Informed Consent
   2. Assessment of Risk and Benefits
   3. Selection of Subjects

Issues related to these guidelines include the protection of “vulnerable subjects,” recruiting volunteers, payment for volunteers, and so on. To ensure that researchers follow these guidelines, research facilities have Institutional Review Boards (IRBs) that must approve all study design plans. If a facility does not have its own IRB, it must contract with an outside source.

The U.S. Department of Health and Human Services’ Food and Drug Administration performs inspections of IRBs to ensure that they are working effectively. (from Food and Drug Administration document at www.fda.gov/oc/ohrt/irbs/reviewboard.pdf)
You will explore the guidelines further through a class share activity. Discuss the following questions and record your responses.

1. Describe why special classes of human subjects need special protection. Select your choices from the following list: fetuses, women, children/minors, cognitively impaired persons, prisoners, traumatized or comatose patients, terminally ill patients, elderly/aged persons, minorities, students, or employees.

2. Could excluding classes of people from research studies be dangerous?

3. How could a survey or questionnaire be dangerous to participants?

4. Why is informed consent of human subjects important to a researcher?

5. What factors should a researcher consider in deciding whether to pay volunteers to participate in a study?

6. REFLECTION: How well do you think the 1974 National Research Act protects the rights and privacy of individuals? In what ways does it support or interfere with the goals of researchers in answering important questions?

7. EXTENSION: Find out about the Institutional Review Board at a research university in your area or online. Do the IRB’s guidelines reflect the provisions of law? Do they go further? Interview a researcher to get his or her perspective on the IRB’s provisions.
Recall the study from Student Activity Sheet 1 (Question 19) that analyzed scalp hair samples from 22 participants with epilepsy and 23 participants without epilepsy, checking for differences in levels of copper, iron, zinc, magnesium, and calcium. The results are summarized in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Copper</th>
<th>Iron</th>
<th>Zinc</th>
<th>Magnesium</th>
<th>Calcium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males, epileptic</td>
<td>14</td>
<td>6</td>
<td>211</td>
<td>329</td>
<td>947</td>
</tr>
<tr>
<td>Males, nonepileptic</td>
<td>19</td>
<td>9</td>
<td>200</td>
<td>259</td>
<td>960</td>
</tr>
<tr>
<td>Females, epileptic</td>
<td>10</td>
<td>7</td>
<td>218</td>
<td>444</td>
<td>1,143</td>
</tr>
<tr>
<td>Females, nonepileptic</td>
<td>16</td>
<td>15</td>
<td>218</td>
<td>505</td>
<td>1,162</td>
</tr>
</tbody>
</table>

Average trace element concentrations (µg/g) in scalp hair

1. If it were possible to measure the presence of copper in the hair of all males with epilepsy in the world, do you think the average would be exactly 14 µg/g? Explain your thinking.

2. The journal article that contains the results of the study actually reports that males with epilepsy have an average of 14 ± 9 µg/g of copper in their scalp hair. What do you think 14 ± 9 means in this situation?

3. The ± 9 is called the *margin of error*. This wording, however, does not mean that someone messed up the research. It simply means that no sampling method can guarantee that the sample exactly matches the population, but that the sampling techniques (when used correctly) can be trusted to give results that are accurate within a certain range.

Since the males with epilepsy in the sample showed an average of 14 µg/g of copper in their scalp hair, the researchers are fairly confident that the true average copper concentration for all males with epilepsy in the study is between 5 and 23 µg/g.

Have you ever seen a news report that mentions *margin of error*? What was the report about?
Politician Paul and Candidate Carl are running for governor, and the election is next week. The latest poll shows that Politician Paul has 46% of the vote, while Candidate Carl has 43% of the vote. The news report, however, states that this poll contains a 3% margin of error.

4. What does this mean for Politician Paul?

5. What does this mean for Candidate Carl?

6. What do these poll results tell you about the upcoming election?

Recall the study from Student Activity Sheet 1 (Question 19) that tested the effect of replacing rabbits’ soybean diet with *Gliricidia sepium* Leaf Meal (GLM). The rabbits were randomly assigned to receive either 0%, 5%, 10%, 15%, or 20% GLM. The resulting effect on weight gain is summarized in the table below.

<table>
<thead>
<tr>
<th>GLM</th>
<th>True Mean Weight Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
</tr>
</tbody>
</table>

Increase in weight during an eight-week period, measured in grams

7. Fill in the interval of the true mean weight gain for each treatment. Based on these results alone, what do you recommend to the farmers in the area? Why?
8. **REFLECTION:** Remember that increasing the concentration of GLM in the rabbit food decreases the cost of the food. Does this change your recommendation? Why or why not?

9. **EXTENSION:** Review an article about not trusting pollsters’ interpretations and the need to see the actual questions. Write a short summary of the article. The following is one such article:

Recall the Spud Potato Chips scenario from Student Activity Sheet 1. You hypothesized that the true mean weight of bags of Spud’s might be less than the 28.3 grams advertised on the bags. Discuss and make some notes on how you might collect a sample of bags to test your hypothesis. Remember that the sample should be representative of the population. What do you mean by “the population of Spud Potato Chips” that you are interested in testing?

This leads to another circular relationship. You want to know something about a population. You choose a sampling method and obtain a sample. Collecting data from the sample provides some sort of estimate or conclusion about the population of interest.

It may seem that the hard work in statistical investigations comes during data collection. Notice, however, that the other four boxes in the research cycle are the same size as the Collect box—they require the same time and attention as the actual data collection.

In particular, researchers cannot skimp on the time and care given to planning sample selection. If the sample is not representative of your population, the results may be worthless!
Recall two of the teachers from Student Activity Sheet 1 (Question 13) who were interested in the best test format for student achievement. Mrs. Johnson flipped a coin to decide whether a student would take a multiple-choice or fill-in-the blank exam. Mr. McDonald asked students which format they preferred.

1. These teachers chose to study the population of all students in their classes. This approach is called a census. The U.S. government conducts a census every 10 years. List some things you know about the U.S. Census.

2. Rather than go through the time and expense of a census, researchers usually choose to sample the population. There are a variety of sampling techniques. Define the following common techniques with help from your teacher or by researching other resources.
   a. Simple random sampling
   b. Stratified random sampling
   c. Systematic sampling
   d. Cluster sampling
   e. Convenience sampling

3. Suppose the school board wished to see whether the age of the student affects test achievement. The testing coordinator separated the roster of high school students into freshmen, sophomores, juniors, and seniors and randomly selected 20 students from each classification. She then flipped a coin to determine which test format each student would receive and then compared the results as follows:
   • freshmen who took the multiple-choice test compared to freshmen who took the fill-in-the blank exam,
   • multiple-choice sophomores compared to fill-in-the-blank sophomores,
   • multiple-choice juniors compared to fill-in-the-blank juniors, and
   • multiple-choice seniors compared to fill-in-the-blank seniors.

Randomization still occurred because the testing coordinator flipped the coin to assign the test format to each student. This approach is an example of what type of sampling technique? Explain your thinking.
4. Mr. McDonald expanded his study to the entire school. He collected a student roster from the office and used the random number generator on his calculator to select one of the first 50 students on the list. Mr. McDonald then selected every 50th student on the list after this initial student for his sample. The calculator generated the number 32. Which students on the roster are the first five in his sample? What type of sampling technique is this? Explain your thinking.

5. Coach Smith wants to know whether students would pay for the privilege of parking their cars in the lot closest to the school. He surveyed students getting on the buses while he monitored bus loading each afternoon. What type of sampling technique is this? What do you think of his plan?

6. A large university wants to find out whether it is adequately serving the needs of its students who live off campus. The campus is surrounded by a large number of apartment complexes. The researchers randomly selected three of the complexes that seemed to contain a diverse group of residents who adequately reflect the student body as a whole, and they surveyed these residents about campus services. What type of sampling technique is this? Why do you think the university chose this method? Explain your thinking.

7. Recall the research from Student Activity Sheet 1 (Question 19) in which scientists analyzed the scalp hair samples from 22 participants with epilepsy and 23 participants without epilepsy, checking for differences in levels of copper, iron, zinc, magnesium, and calcium. The scientists were concerned about previous research that showed conflicting results. They speculated that other differences in the study group, besides the presence or absence of epilepsy, could have caused these mixed results. The scientists attempted to control some of these other differences by gathering all participants from the same region of Turkey (indicating similar dietary habits) and separating the participants into the following groups:
   - males with epilepsy,
   - males without epilepsy,
   - females with epilepsy, and
   - females without epilepsy.

This is an example of what kind of sampling? Explain your thinking.
8. Recall the study from Student Activity Sheet 1 (Question 19) that tested the effect of replacing rabbits’ soybean diet with *Gliricidia sepium* Leaf Meal (GLM). Twenty-five young rabbits were randomly assigned to receive either 0%, 5%, 10%, 15%, or 20% GLM. Suppose, rather than random assignment, the scientists chose the following method:

   The research assistant who was in charge of gathering rabbits went in the barn and assigned the first five rabbits he could catch to the 0% group. He assigned the next five that he caught to the 5% group, and so on.

   What type of sampling technique is this? What do you think of his plan?

9. **REFLECTION:** During busy political seasons, many opinion polls are conducted. In a presidential race, how do you think the participants in polls are generally selected? Discuss any issues regarding simple random, stratified, systematic, cluster, and convenience sampling in these polls. What about other types of polls, besides political?

10. Describe briefly how each technique could be used in the potato chip investigation. Which techniques are the most appropriate?

   - Simple random sampling
   - Stratified random sampling
   - Systematic sampling
   - Cluster sampling
   - Convenience sampling
11. The most feasible method is to ask other U.S. schools to help with cluster sampling. On a smaller scale, you could mimic this by cluster sampling from grocery stores in the area. The easiest choice (but least reliable) is convenience sampling.

Pursue the idea of cluster sampling with other schools. To do this, you simply modify random selection techniques to choose your schools. Following are two ways to use random numbers for selection.

**Method 1: Random Number Tables**

Such tables can be found on the Internet by typing “random number tables” in the search box. Here is an example of what you might get:

<table>
<thead>
<tr>
<th>241983</th>
<th>724152</th>
<th>579108</th>
<th>492124</th>
<th>912127</th>
<th>508114</th>
<th>280505</th>
<th>344304</th>
</tr>
</thead>
<tbody>
<tr>
<td>298477</td>
<td>911677</td>
<td>859342</td>
<td>730503</td>
<td>184740</td>
<td>934279</td>
<td>233161</td>
<td>887766</td>
</tr>
<tr>
<td>847111</td>
<td>724875</td>
<td>393074</td>
<td>591162</td>
<td>996737</td>
<td>358072</td>
<td>852052</td>
<td>761457</td>
</tr>
<tr>
<td>330180</td>
<td>678434</td>
<td>267867</td>
<td>657965</td>
<td>812675</td>
<td>230136</td>
<td>276862</td>
<td>466559</td>
</tr>
<tr>
<td>711671</td>
<td>103110</td>
<td>259433</td>
<td>112082</td>
<td>050556</td>
<td>058988</td>
<td>273557</td>
<td>154354</td>
</tr>
<tr>
<td>533800</td>
<td>188724</td>
<td>706462</td>
<td>730447</td>
<td>481964</td>
<td>913634</td>
<td>627872</td>
<td>459415</td>
</tr>
<tr>
<td>564428</td>
<td>022187</td>
<td>445787</td>
<td>920698</td>
<td>352175</td>
<td>172443</td>
<td>463765</td>
<td>576652</td>
</tr>
<tr>
<td>981420</td>
<td>452204</td>
<td>121938</td>
<td>693647</td>
<td>353129</td>
<td>156861</td>
<td>696173</td>
<td>707475</td>
</tr>
</tbody>
</table>

If you have 100 schools and wish to select 12 of them to help with the study, you assign each school a number between 00 and 99. Then read the table from left to right and select the schools as their numbers appear. The 12 schools are Nos. 24, 19, 83, 72, 41, 52, 57, 91, 8, 49, 21, and 27.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>241983</td>
<td>724152</td>
<td>579108</td>
<td>492124</td>
<td>912127</td>
<td>508114</td>
<td>280505</td>
<td>344304</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>298477</td>
<td>911677</td>
<td>859342</td>
<td>730503</td>
<td>184740</td>
<td>934279</td>
<td>233161</td>
<td>887766</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>847111</td>
<td>724875</td>
<td>393074</td>
<td>591162</td>
<td>996737</td>
<td>358072</td>
<td>852052</td>
<td>761457</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>330180</td>
<td>678434</td>
<td>267867</td>
<td>657965</td>
<td>812675</td>
<td>230136</td>
<td>276862</td>
<td>466559</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. If you have 1,000 schools on the list, number them from 000 to 999 and read the table three digits at a time. What are the first five schools in this sample?

b. To avoid using the same numbers for every study, you can start from a different row. If you read three digits at a time starting with the third row, what five schools do you put in your sample?
Method 2: Random Number Generators
The same task can be accomplished with your graphing calculator or a random number generator found on the Internet.

c. If you use the calculator to choose a sample from the list of 1,000 schools, what are the first five schools in your sample? (Hint: Remember that you need three-digit numbers.)

12. EXTENSION

Dream Car
Students’ tastes vary by grade level
Do you remember what car you liked best when you were in 7th grade? What car do you like best now? Are the cars different because your tastes have matured or because new car styles have been introduced?

To investigate whether students’ tastes vary by grade level, students from every grade level (7-12) must be represented fairly. You decide to use a sample size of 60. Since there are six grade levels, you could randomly select 10 students from each grade. Notice, however, that the 8th-grade class is much larger than the 12th-grade class. Should both of these grade levels have the same number of students in the sample? One way to fairly represent each grade is to use a stratified sample. Following are sample enrollment numbers.

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>No. of Students</th>
<th>Percentage of Total</th>
<th>No. of Students for Sample</th>
<th>No. of Students for Sample (Rounded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7th</td>
<td>476</td>
<td>17.539</td>
<td>10.52</td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>511</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9th</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10th</td>
<td>473</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11th</td>
<td>425</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12th</td>
<td>337</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2,714</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Student: ____________________________  Class: ____________________________  Date: ________________

**Statistical Studies: Statistical Investigations**  
III.A Student Activity Sheet 4: Sampling Design and Methods

**a.** Seventh-graders represent \(\frac{476}{2,714}\), or 17.539% of the total. Fill in the table column showing the percentage of the total represented by each grade.

**b.** Since you plan to collect data from a sample size of 60, fill in the fourth column of the table indicating how many of the sample should be allocated to each grade. For example, the 7th graders get \(0.17539 \times 60 = 10.52\) representatives in the sample.

**c.** In the fifth column of the table, round each number in the normal way. What sample size does this give?

**d.** How should you award the extra person(s)?

**13. EXTENSION**

Researchers must disclose their methods and any potential problems to readers of their study. Researchers can also make recommendations for future studies that improve on their methods.

Look at the article, “Are Women Really More Talkative Than Men?” and answer the following questions. There are two different studies discussed in the article, so some questions may have two answers. You may have to infer some answers from the reading.

**a.** What was the population of interest?

**b.** How were the samples obtained?

**c.** What were some problems that researchers noted about their study designs?

**d.** What were some recommendations for future study?
14. EXTENSION

Cavities and Kids
Does eating candy increase cavities?

Have you ever seen a news report on the effect that eating candy has on children’s teeth? To study this subject further, what are some issues you would have to consider?

a. Describe an observational design to study this issue. Be sure to include a description of the population of interest, your sample selection technique, and what variables will be measured.

b. Describe an experimental design to study this issue. Be sure to include a description of the population of interest, your sample selection technique, and what variables will be measured.

c. What is a key difference in what can be reported for each design type?

d. Are there any other issues to consider in this study? For example, do you want to put all children in a group together? Or can you think of a way in which stratified sampling is more appropriate?

15. EXTENSION

“When I Grow Up…”
What careers are of most interest to teens?

a. What do you think a news report on teens’ career plans would say?

b. If you wanted to study this subject further, what are some issues you must consider?

b. Write three questions that could address those issues.
1. What does the histogram in the opening of the lesson represent? Describe the distribution as completely and accurately as possible with regard to center, shape, and spread.

2. Refer to the research cycle. Where in the cycle is examining histograms located?
Statistical Studies: Analyzing Data
III.B Student Activity Sheet 5: Histograms

So, let’s collect some data!

3. Without discussing with your classmates, answer the following questions—write one answer on each of your 12 slips of paper. Place each slip in the appropriate paper bag.
   a. What year were you born?
   b. What is your gender?
   c. How many text messages did you send yesterday?
   d. What does your cell plan charge for texting?
   e. How many people under the age of 18 live in your house?
   f. Which is your favorite food from the following choices: pizza, hamburgers, sushi, salad, chicken, other?
   g. Do you have a job that pays by the hour?
   h. If yes, how many hours do you work in an average week? (If no, put 0.)
   i. List how many hours you work a week (again) and then the average number of hours that you study a week.
   j. List your gender (again) and your shoe size.
   k. List your gender (again) and number of text messages (again).
   l. What is your favorite kind of music?

4. While your classmates finish Question 3, jot down your thoughts about the following questions:
   • What do you think categorical data means?
   • What do you think quantitative data means?
   • What do you think univariate means?
   • What do you think bivariate means?
5. Does the histogram illustrate univariate data or bivariate data? Categorical or quantitative?

6. Would you describe the histogram as symmetric? Why or why not?

7. Describe the distribution of grades for the class. Justify any estimates you make.
8. Consider the new histogram showing bus ridership at a local high school. This histogram is not symmetric; it is skewed to the right. You know this because the distribution has a tail out toward the right side. What is happening with this population that causes the distribution to be skewed to the right?

9. Estimate the average age of students who ride the bus. This average is the center of the distribution. Justify your estimate.

10. Suppose the histogram of bus ridership looked like this instead. Data values that are distant from most of the other values are generally thought of as outliers. What may have happened? Does this data distribution affect your answers to Questions 8 and 9?
11. **REFLECTION:** Make a histogram that represents the number of students of each age at the same school who drive to school each day (either in their own car or riding with friends). Describe the center, shape, and spread of your histogram.

Several of the previous histograms, including the one in Question 10, had a set interval width (for example, the ages of bus riders in intervals of one year). Now you will investigate the effect that interval width (also known as **bin size**) can have on a histogram as well as on a reader of the histogram.

Go to the following website and select the “Colleges’ SAT Math Scores” data set:

illuminations.nctm.org/ActivityDetail.aspx?id=78


12. The default interval width for this data set is 19.1. Use the slider under the histogram to change the interval width to approximately 30. What does this do to the appearance of the graph? Explain.

13. Change the interval width to approximately 50. Comment on the new graph.

14. Try an interval of approximately 75. Interpret your result.

15. Try an interval of 250. What do you think of this graph and the information it communicates?

16. Reset the interval size back to approximately 75. Change the value on the Maximum Frequency on Histogram box to 300 (click the Update Histogram button to enact the change). What do you think of this change to the graph and the information it communicates?

17. **EXTENSION:** Above the histogram is a drop-down menu from which you can choose another data set. Select one of the other data sets and explore a variety of interval widths for the graph. Sketch three histograms that show how changing the interval width results in widely differing perceptions for the reader of the graph. Explain each choice.
18. **EXTENSION:** Use the drop-down menu to choose a third data set. Once again, sketch three histograms that show how changing the interval width results in widely differing perceptions for the reader of the graph. Explain each choice.

19. **EXTENSION:** Use the drop-down menu to choose My Data. Go to the Describe Your Data box below the histogram and give your histogram a name. Below that is a data input box (*Enter your data below, one per line*). Enter your data and try a variety of interval widths. Choose two graphs to sketch—one that accurately communicates the data to the reader and another that distorts the data. Explain each choice.
The Phoenix Mercury of the Women’s National Basketball League had 14 players on the roster for the 2008 season. The players and their average points per game (PPG) are shown below.

<table>
<thead>
<tr>
<th>Player</th>
<th>Diana Taurasi</th>
<th>Cappie Poindexter</th>
<th>Tangela Smith</th>
<th>Le’coe Willingham</th>
<th>Kelly Miller</th>
<th>Kelly Mazzante</th>
<th>LaToya Pringle</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPG</td>
<td>24.1</td>
<td>21.2</td>
<td>11.1</td>
<td>10.1</td>
<td>8.3</td>
<td>5.8</td>
<td>4.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Player</th>
<th>Brooke Smith</th>
<th>Barbara Farris</th>
<th>Olympia Scott</th>
<th>Yuko Oga</th>
<th>Allie Quigley</th>
<th>Willnett Crockett</th>
<th>Jennifer Derevjanik</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPG</td>
<td>4.1</td>
<td>3.5</td>
<td>2.7</td>
<td>2.4</td>
<td>2.1</td>
<td>1.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>

(Source: www.wnba.com/mercury/stats)

1. The smallest value listed in a data set is called the **minimum**. The minimum of this data set is ________. Which player has the minimum value?

2. The largest value listed in a data set is called the **maximum**. The maximum of this data set is ________. Which player has the maximum value?

3. The middle value in a data set is called the **median**. The median of this data set is ________. Which player has the median value? (Note: This problem requires some work because a middle value does not exist when there is an even number of data points. In this case, you must average together the two middle values.)

4. Now list the data set horizontally from smallest to largest, and write the median in the list in the appropriate location. Circle the minimum, median, and maximum.

5. Cover up the right side of the list in Question 4 so that you can only see the seven values below your median. Find the median of these seven numbers and circle it; under that number write **Q1**. Repeat this process with the other half of the data by covering up the left side so that you can only see the seven highest values. Find the median of these seven numbers and circle it; under that number write **Q3**.
The numbers you have circled are called the **five-number summary**. These numbers separate your data into four quartiles, or 25% sections.

- The data between the minimum and Q1 are the first quartile.
- The data between Q1 and the median are the second quartile.
- The data between the median and Q3 are the third quartile.
- The data between Q3 and the maximum are the fourth quartile.

6. The five-number summary allows you to make a graphical display called a **boxplot**, or a **box-and-whisker plot**. The reason for this interesting name becomes obvious as you construct the graph. First you need to decide on a scale. What would be a good scale for these data—to count by 1s, 10s, 100s, or something in between?

7. Construct a box-and-whisker plot. The following steps are provided for your reference.
   - Plot your scale on the line below.
   - Place an appropriate label below the line.
   - Place dots for your five-number summary values about an inch above the line.
   - Put a small vertical line, about the size of this $l$, on each dot.
   - Use these lines to construct a box-and-whiskers like this one:

8. Interpret the “box” part of your box-and-whisker plot:

   50% of the Phoenix Mercury players ________________________________.
9. You can also create boxplots on your graphing calculator. Sketch your new graph—compare and contrast it to your hand-drawn boxplot.

   There may also be an option on your calculator for creating a modified box plot, which reveals any outliers. If so, sketch this graph. What do you think this graph is showing that is different from your previous one? (If you have access to this information using technology, research information about this plot that distinguishes it from a box-and-whisker plot.)

10. EXTENSION: After you finish the hand-constructed and calculator graphs, look up statistics for another WNBA team (or NBA team), compute the five-number summary, and add it to the same graph from earlier teams. Thus, you have created a side-by-side boxplot. Compare and contrast the two boxplots.

Take a look at other graphical displays. Consider the characteristics of the center, shape, spread, and any unusual features.

11. The school newspaper conducted a survey in which 31 randomly selected students were asked a variety of questions. The responses to one question are shown in the following dotplot (lineplot). Discuss what you now know about these students.

     X
     X
     X
     X
     X  X  X
     X  X  X  X  X
     X  X  X  X  X  X  X
     X  X  X  X  X  X  X  X
     X  X  X  X  X  X  X  X

     0 1 2 3 4 5 6 7 8 9 10 11 12

     Number of hours of sleep the previous night
12. Some members of the newspaper staff wanted to report the sleep data in a frequency table as shown below. Discuss the advantages and disadvantages of this option.

<table>
<thead>
<tr>
<th>Number of hours of sleep</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>

13. Other staff members voted for a boxplot. Compare and contrast the usefulness of this boxplot with that of the previous two graphical displays.

```
Number of hours of sleep the previous night
```

```
 0 1 2 3 4 5 6 7 8 9 10 11 12
```

---

Charles A. Dana Center at The University of Texas at Austin
Advanced Mathematical Decision Making (2010)
Activity Sheet 6, 6 pages
14. Another group of staffers argued for the following graph. What were their reasons for preferring a histogram? What are the arguments against using one?

![Histogram](image)

**Number of hours of sleep the previous night**

15. The following dotplots show the effect of separating the data on male students' hours of sleep from the data on female students' hours of sleep. Compare and contrast the two plots.

- Female dotplot:

  ![Female Dotplot](image)

  **Number of hours of sleep the previous night - females**

- Male dotplot:

  ![Male Dotplot](image)

  **Number of hours of sleep the previous night - males**
16. Choose the frequency table, boxplot, or histogram format to report the male and female sleep data.

17. REFLECTION: Is it easier to compare and contrast the male and female sleep data from the dotplot or from your new display in Question 16? Explain. Refer to the information regarding limitations and differences of various graphical representations to support your thinking as needed.

18. EXTENSION: For one of the sets of data in Question 3 in Student Activity Sheet 5, represent the data in as many different graphical displays as possible.

19. EXTENSION: Write two reports—an informative paper on student sleep patterns (or other sets of data of their choice) and a persuasive paper that utilizes the data and graph.
Suppose data were collected on 25 bags of Spud Potato Chips. The weight (to the nearest gram) of the chips in each bag is listed below.

<table>
<thead>
<tr>
<th>25</th>
<th>28</th>
<th>23</th>
<th>26</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>24</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>23</td>
<td>24</td>
<td>28</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>26</td>
<td>24</td>
<td>25</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>25</td>
<td>26</td>
<td>24</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

1. Create a dotplot of the potato chip data and describe the distribution.

2. Does this distribution appear to support or contradict the manufacturer’s claim of an average weight of 28.3 grams of chips per bag? Explain your reasoning.

3. Use your calculator or spreadsheet software to create a more formal display of these data. Make a sketch of the result here and save an electronic copy for your formal report. (If you create a histogram, remember to carefully consider your bin size.)

4. You can also use your calculator or spreadsheet software to create more precise numerical descriptions of the data. This approach is quicker and usually more accurate than computing by hand.

   The mean of the data is \( \bar{x} = \)

   The standard deviation of the data (a measure of “spread-outness”) is \( S_x = \)

   The number of data values is \( n = \)

   The five-number summary is

   The numbers in the five-number summary represent

5. Sometimes it makes sense to analyze the proportion of a population that meets some criterion. This is another method for investigating whether the manufacturer is correct in claiming that the average weight of a bag of Spud’s is 28.3 grams. Write a statement or hypotheses involving the proportion of bags of chips that meet a weight criterion.
6. What proportion of the bags in your sample were 28.3 grams or more? Does the answer change your opinion about Question 2? Why or why not?

7. Suppose instead that the weight of the chips in each Spud’s bag is the following:

```
29  34  22  27  26
25  28  24  26  33
28  29  31  30  27
28  31  28  32  25
31  28  30  29  27
```

Produce a histogram and recompute the descriptive statistics for this set of data. What do you notice?

The mean of the data is $\bar{x} =$

The standard deviation of the data is $S_x =$

The number of data values is $n =$

The five-number summary is

8. **EXTENSION:** Prepare a professional report for the president of Spud Potato Chips based on one of the data sets.
Data Set 1

Go to www.billboard.com and select the Hot 100. This list gives the top 100 songs for the week. You will compute statistics on the number of weeks that a random selection of the songs has been on the charts. (If you wish, choose one of the specialized charts such as R&B/Hip-Hop or Country.)

Use your calculator or a random number table to select 10 of the 100 songs. Write down the name of each song, its rank on the list, and the number of weeks it has been on the chart. For example, if your random number generator gives you 3, write “No. 3, Mary Had a Little Lamb, 14 weeks.”

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Data Set 2

Go to www.imdb.com and select IMDb Top 250. (To find this link, scroll down and look for User’s Favorites in the sidebar on the left.) This list ranks the top 250 movies as voted by regular users of the Internet Movie Database (IMDb). (Note: These are listed by ranking, not by number of voters. The voters did not vote for a favorite movie; rather, they ranked the movies on a scale of 0-10). You will compute statistics on the number of voters who provided input on the movies, without actually using 250 pieces of data in your calculation. (If you wish, choose the IMDb Bottom 100 instead.)

Use your calculator or a random number table to select 10 of the 250 movies. Write down the name of each movie, its rank on the list, and the number of voters who ranked it.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?
Data Set 3
For information about state sales taxes, go to the Federation of Tax Administrators’ website at www.taxadmin.org/FTA/rate/sales.html. The sales tax for each state and the District of Columbia is listed. You will compute statistics for state sales taxes, without actually using 51 pieces of data in your calculation.

Use your calculator or a random number table to select 10 of the 51 locations. Write down the name of each location and its sales tax.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Data Set 4
To view gasoline sales taxes by state, go to the American Petroleum Institute website at www.api.org/statistics/fueltaxes. You will compute statistics for gasoline sales taxes by state, without actually using 51 pieces of data in your calculation.

Use your calculator or a random number table to select 10 of the 51 amounts. Write down the name of each state and its sales tax on a gallon of gas.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?
Data Set 5
To view diesel sales taxes by state, go to the American Petroleum Institute website at www.api.org/statistics/fueltaxes. (Scroll down to find the diesel data, which appear after the gasoline data). You will compute statistics for diesel sales taxes by state without, actually using 51 pieces of data in your calculation.

Use your calculator or a random number table to select 10 of the 51 amounts. Write down the name of each state and its sales tax on a gallon of diesel fuel.

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?

Data Set 6
To view the tournament records for men’s college basketball teams, go to webpages.charter.net/dbwoerner/coaches/schl109.htm. Because there are hundreds of schools listed, you will take a sample to compute the average winning percentage.

Use your calculator or a random number table to select 25 teams. Write down the name of each school and compute the winning percentage \( \frac{\text{wins}}{\text{total games played}} \).

Calculate the descriptive statistics for the data set and interpret all statistics.

Write a couple of sentences for the school newspaper about your results.

What would a graph tell you about these data?
Statistical Studies: Analyzing Data
III.B Student Activity Sheet 7: Using Technology

Data Set 7
Visit the website for the National Assessment of Educational Progress (NAEP) at nces.ed.gov/NATIONSREPORTCARD/states. Click on State Comparisons. Circle the subject that your teacher assigns to you:

Mathematics  Reading  Science  Writing

Choose Grade 8, your assigned subject, Gender, and the most recent testing year. Then click on the Next Steps button; this generates a data table from the 50 states, the District of Columbia, and Department of Defense schools. Circle the data display that your teacher assigns to you:

Histogram  Boxplot

Compute the appropriate statistics for each gender’s data set.

Create a comparative graphical display.

Write a technical report comparing and contrasting the data that might be submitted to the NAEP website for publication.

Prepare a nontechnical version of your report to be presented to students and parents.
Data Set 8
Visit the website for the U.S. Census Bureau at www.census.gov. In the People & Households category, click Estimates, then Estimates Data, and then Totals. Under Vintage 2009, click Annual Population Estimates. You can view the Excel document with population estimates for each state. Make a list of the 10 least populous states and a list of the 10 most populous states.

Next, circle the website that your teacher assigns to you:

- the American Petroleum Institute at www.api.org/statistics/fueltaxes to view gasoline sales taxes by state
- the Census Bureau at www.census.gov to view household incomes; in the People & Households category, choose State Median Income
- Swivel at www.swivel.com/data_sets/spreadsheet/1006019 to view the number of cars per state
- Swivel at www.swivel.com/data_sets/spreadsheet/1000483 to view the crime rate per state

Go to your assigned website and collect data for the 20 states on your list. You will display these data in one of two ways. Circle the data display that your teacher assigns to you:

- Histogram
- Boxplot

Compute the appropriate statistics for your assigned data set.

Create a comparative graphical display.

Write a technical report comparing and contrasting the data that might be submitted to the Census Bureau for publication.

Prepare a nontechnical version of your report to be presented to the local media.
Statistical Studies: Analyzing Data
III.B Student Activity Sheet 7: Using Technology

Data Set 9
Visit the website for the National Collegiate Athletic Association (NCAA) at www.ncaa.org/wps/portal. Go to Statistics & Records and choose a sport.
Which sport did you choose?

Click on Archived Team-by-Team Final Statistics. (Note: This may be worded slightly differently, depending on the sport you choose.)

Choose two different schools. (Note: Remember that schools do not play every sport, so if you do not get results, pick another school). Which two schools did you choose?

Choose a category such as Points Per Game or Assists and record the data for the players at each school.

Circle the data display that your teacher assigns to you:

Histogram          Boxplot

Compute the appropriate statistics for each data set.

Create a comparative graphical display.

Write a technical report comparing and contrasting the data that might be submitted to the NCAA for publication.

Prepare a nontechnical version of your report to be presented to local sportswriters.
“School uniforms should not be required.”
Agree or disagree?

“Standardized dress can promote a productive school environment.”
Agree or disagree?

These questions show that the design of a survey can influence the results. Questions should be designed to be neutral and to allow the accurate recording of the opinions or facts given by the participants. For example, the uniform question could be worded as follows:

*I am in favor of school uniforms for high school students.*

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>No opinion</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
</table>

This is an example of a *closed question* because the participants are limited to the response choices that are given. Closed questions are easier (and usually cheaper) to analyze, and sometimes a computer does much of the work. *Open questions* allow the participants to give more detailed responses. This approach, however, requires a follow-up analysis that is more labor intensive (and more expensive).

1. Reword the school uniform question so that it is an open question.

Because the design of a survey is so important for capturing the information needed accurately, *pilot surveys* are often conducted. A pilot survey is used with a small number of people and then analyzed to look for the following:

- question ambiguity,
- leading questions,
- too many people choosing Other or No opinion, and
- other data collection problems.
2. The following questions can be worded more effectively. Describe the question’s problem and what effect the problem could have on the results. Then come up with more effective wording for the question. Provide a closed question and an open question.

   a. I can always talk to my parents about my problems.                  True/False
   b. Teachers and students like the new school schedule.                True/False
   c. I like the school cafeteria.                                      True/False
   d. I support school activities.                                       True/False

When designing a survey (or an observation instrument), the researcher must also consider how the data will be collected. The following are some options:

   a. A face-to-face interviewer asks questions and records the answers of the participants (on paper or an electronic recording).
   b. A telephone interviewer asks questions and records the answers.
   c. A computer calls homes and records answers through key presses.
   d. A researcher observes behaviors or characteristics and marks the survey.
   e. A researcher hand-delivers surveys to participants, who then fill out the surveys and turn them back in.
   f. The survey is mailed to participants.
   g. The survey is e-mailed to participants.

3. Discuss the pros and cons of each method with your partner(s). Record your observations.

Always bear in mind that whatever method of data collection is used, it is vitally important that the results be accurately interpreted and reported. This includes thoroughly disclosing the methods to the reader.
Recall Spud Potato Chips one last time. One sample that was “collected” had the results shown below. If the true mean of all Spud’s bags is really 28.3 grams, as claimed by the company, this sample appears to have problems. The sample mean is far below the true mean and, in fact, all the bags are below the true mean.

\[
\begin{align*}
\text{True population mean} &= 28.3 \\
\bar{x} &= 25.16 \\
S_x &= 1.46
\end{align*}
\]

This discrepancy between a population parameter and a sample statistic is known as **statistical bias**, which can result from many different sources. Two broad categories of statistical bias are **biased sampling method** and **biased statistic**.

1. Give an example of a biased sampling method.

2. Sometimes researchers spend vast resources (time, money, effort) to get a great sample and then still end up with a biased statistic. For example, something could be wrong with the data collection method. What could happen after selecting the sample (for example, potato chip bags) and before calculating the statistic that could result in a biased statistic?
If in fact the method of sampling has introduced bias, one reason for this might be that the sample was actually nonrepresentative of the population (nonrepresentative sampling).

Nonrepresentative sampling is when the sample does not represent the differences in a population. For example, the students in this class are mainly seniors in high school; therefore, they do not represent all students in the school. If only the football team or only the volleyball team were surveyed at most schools, the sample would have all males or all females.

3. What other situations could be responsible for bias? Provide examples where possible.

4. REFLECTION: Describe a scenario in which different types of sampling methods lead to different kinds of bias in sampling, including yielding nonrepresentative samples or samples with undercoverage of a population.

5. Biased statistics can result from a variety of problems during the data collection process. Record your thoughts or information you have located through research regarding each of the following sources of statistical bias.
   a. Response bias
   b. Nonresponse bias
   c. Observer effect
   d. Wording of questions
   e. Placebo effect
Recall that a second sample of Spud Potato Chips was “collected,” and the following results were obtained:

\[
\begin{array}{cccccccccccc}
\text{22} & \text{23} & \text{24} & \text{25} & \text{26} & \text{27} & \text{28} & \text{29} & \text{30} & \text{31} & \text{32} & \text{33} & \text{34} \\
\hline
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\hline
28.32 & 2.84 \\
\end{array}
\]

6. Comment on this distribution compared to that of the original Spud’s sample.

7. **REFLECTION:** This sample of chips is an example of **high variability** and no statistical bias. This situation can be a big problem for the manufacturer. People who get bags with only 22 grams of chips probably do not really care what the mean is—they just care that they were shorted! Compare the standard deviations that are given for the original distribution on Page 1 and this distribution.

What are your recommendations for the manufacturer?
8. Identify each of the following as high or low statistical bias and high or low variability.

```
[Graphs showing bar charts with population mean indicated]
```

9. REFLECTION: In your opinion, which graph in Question 8 represents the worst situation for researchers? Explain your reasoning.

10. EXTENSION: Variability can be caused by natural variability or induced variability. Research each, record your findings, and provide some examples.
   a. Natural variability
   b. Induced variability
Statistical Studies: Sources of Variability
III.C Student Activity Sheet 10: Statistical Bias in Research Studies and Polls

Statistical bias in experimental studies can result from several factors. Look back at your notes about statistical bias and see which sources you think could apply to experimental studies. List your ideas below.

Experimental studies can suffer from nonrepresentative samples and undercoverage, as discussed in Student Activity Sheet 9. Often these problems cannot be fixed.

1. What are some reasons for deliberately using a nonrepresentative sample in an experimental study? What problems could result?

2. Rather than spending a great deal of time and money to ensure a representative sample, researchers often use the techniques listed below to try to eliminate any statistical bias introduced by the sampling method. Discuss how each method can reduce sampling method bias and thus increase the accuracy of a study’s results.

   a. Random assignment of treatments
   b. Blind/double-blind studies
   c. Use of control groups
   d. Replication
3. Some of the same issues (nonrepresentative samples, undercoverage, replication) affect observational studies. Other issues discussed (response bias, wording of questions) can also specifically affect these studies. What problem could be occurring in each study described?

   a. Jade is so tired of sales calls that she refuses to answer the phone if she does not recognize the name or number on the Caller ID.

   b. Serena knows the caller will probably keep calling, so she answers the phone. If the call’s subject is an opinion poll, she politely tells the surveyor that she does not want to participate.

   c. DeShawn does not worry about sales calls since his family uses cell phones and no longer has a land line.

   d. At the mall, an older man is interviewing people about their consumer spending habits. Maria does not want to tell him she just bought a swimsuit, so she says her shopping bag contains a new shirt.

   e. Mrs. Gibbs normally gets a hamburger for lunch. When she notices that someone who looks like a nurse is sitting in the corner writing down what each person orders, Mrs. Gibbs orders a salad with fat-free dressing.

   f. The local television station runs a viewer poll on the nightly news: “Call us or log on to our website to give your opinion about the new skate park.”

4. Several television shows have a survey component in which viewers call in and vote for their favorite singer/dancer/entertainer/competitor. Discuss the issues you perceive with this method.

5. Some shows have viewers vote for the contestant whom they want to leave the show. Discuss the issues you perceive with this method.

6. REFLECTION: What suggestions do you have for improving call-in or online polls or contests?
7. **EXTENSION:** Discuss with your partner(s) where in the research cycle is statistical bias most likely to occur. Suggest at least two strategies for avoiding bias and minimizing variability in your project.

![Research Cycle Diagram]

- Question
- Design
- Collect
- Analyze
- Report