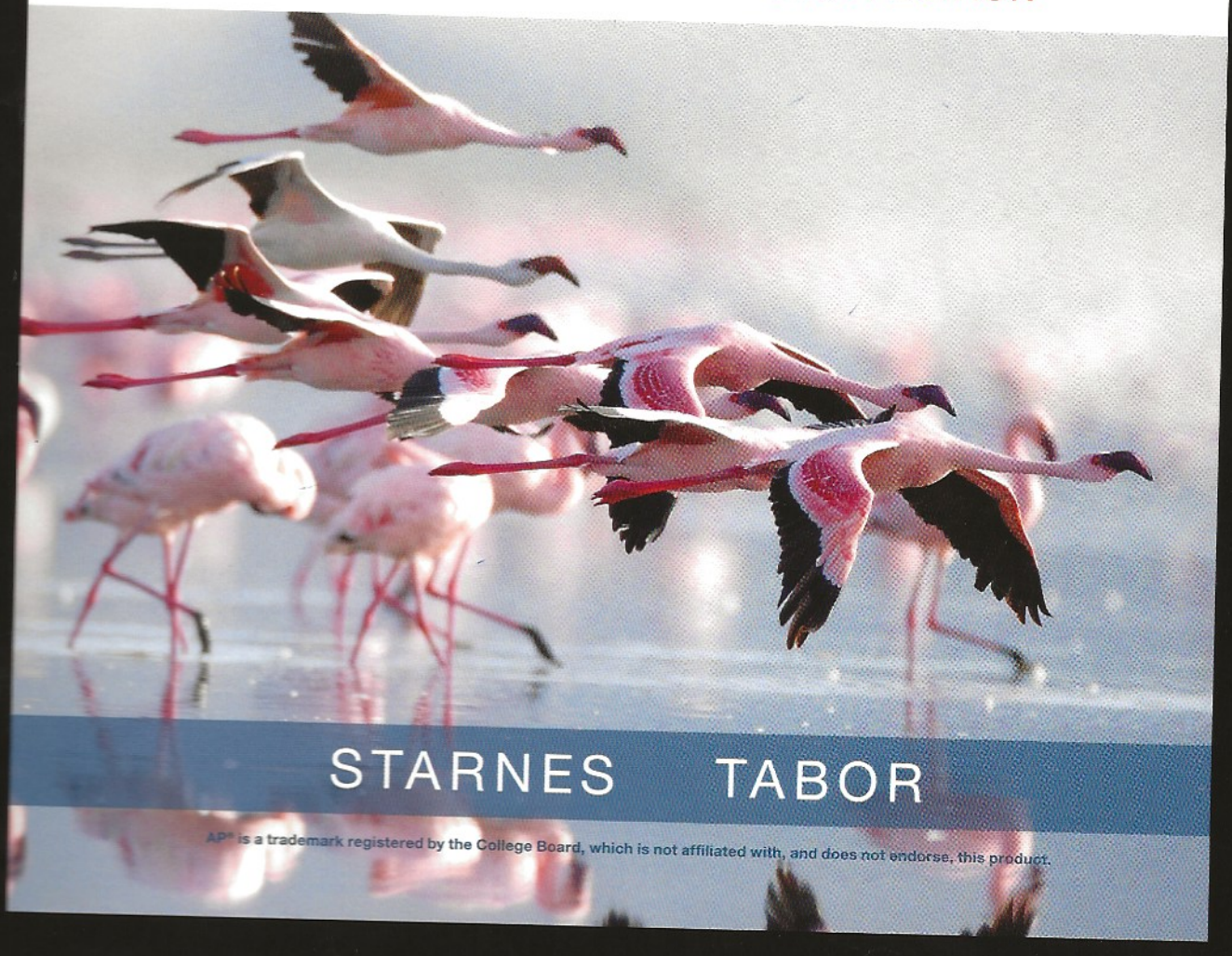
UPDATED

to reflect the revised Course Framework

for the AP® Exam

The Practice of Statistics

SIXTH EDITION



STARNES TABOR

To the Student

Statistical Thinking and You

The purpose of this book is to give you a working knowledge of the big ideas of statistics and of the methods used in solving statistical problems. Because data always come from a real-world context, doing statistics means more than just manipulating data. *The Practice of Statistics* (*TPS*), Sixth Edition, is full of data. Each set of data has some brief background to help you understand where the data come from. We deliberately chose contexts and data sets in the examples and exercises to pique your interest.

TPS 6e is designed to be easy to read and easy to use. This book is written by current high school AP® Statistics teachers, for high school students. We aimed for clear, concise explanations and a conversational approach that would encourage you to read the book. We also tried to enhance both the visual appeal and the book's clear organization in the layout of the pages.

Be sure to take advantage of all that *TPS* 6e has to offer. You can learn a lot by reading the text, but you will develop deeper understanding by doing the Activities and Projects and answering the Check Your Understanding questions along the way. The walkthrough guide on pages xvi–xxii gives you an inside look at the important features of the text.

You learn statistics best by doing statistical problems. This book offers many different types of problems for you to tackle.

- **Section Exercises** include paired odd- and even-numbered problems that test the same skill or concept from that section. There are also some multiple-choice questions to help prepare you for the AP® Statistics exam. Recycle and Review exercises at the end of each exercise set involve material you studied in preceding sections.
- **Chapter Review Exercises** consist of free-response questions aligned to specific learning targets from the chapter. Go through the list of learning targets summarized in the Chapter Review and be sure you can say of each item on the list, "I can do that." Then prove it by solving some problems.
- The **AP® Statistics Practice Test** at the end of each chapter will help you prepare for in-class exams. Each test has about 10 multiple-choice questions and 3 free-response problems, very much in the style of the AP® Statistics exam.
- Finally, the **Cumulative AP® Practice Tests** after Chapters 4, 7, 11, and 12 provide challenging, cumulative multiple-choice and free-response questions like those you might find on a midterm, final, or the AP® Statistics exam.

The main ideas of statistics, like the main ideas of any important subject, took a long time to discover and thus take some time to master. The basic principle of learning them is to be persistent. Once you put it all together, statistics will help you make informed decisions based on data in your daily life.

TPS and AP® Statistics

The Practice of Statistics (*TPS*) was the first book written specifically for the Advanced Placement (AP®) Statistics course. This updated version of *TPS* 6e is organized to closely follow the AP® Statistics Course Framework (CF). Every learning objective and essential knowledge statement in the CF is covered thoroughly in the text. Visit the book's website at highschool.bfwpub.com/updatedtps6e for a detailed alignment guide, including "The Nitty Gritty Alignment" guide to EKs and LOs in Updated *TPS* 6e. The few topics in the book that go beyond the AP® Statistics syllabus are marked with an asterisk (*).

Most importantly, *TPS* 6e is designed to prepare you for the AP® Statistics exam. The author team has been involved in the AP® Statistics program since its early days. We have more than 40 years' combined experience teaching AP® Statistics and grading the AP® exam! Both of us have served as Question Leaders for more than 10 years, helping to write scoring rubrics for free-response questions. Including our Content Advisory Board and Supplements Team (page vii), we have extensive knowledge of how the AP® Statistics exam is developed and scored.

TPS 6e will help you get ready for the AP® Statistics exam throughout the course by:

- **Using terms, notation, formulas, and tables consistent with those found in the Course Framework and on the AP® Statistics exam.** Key terms are shown in bold in the text, and they are defined in the Glossary. Key terms also are cross-referenced in the Index. See page F-1 to find "Formulas for the AP® Statistics Exam," as well as Tables A, B, and C in the back of the book for reference.
- **Following accepted conventions from AP® Statistics exam rubrics when presenting model solutions.** Over the years, the scoring guidelines for free-response questions have become fairly consistent. We kept these guidelines in mind when writing the solutions that appear throughout *TPS* 6e. For example, the four-step State–Plan–Do–Conclude process that we use to complete inference problems in Chapters 8–12 closely matches the four-point AP® scoring rubrics.
- **Including AP® Exam Tips in the margin where appropriate.** We place exam tips in the margins as "on-the-spot" reminders of common mistakes and how to avoid them. These tips are collected and summarized in the About the AP® Exam and AP® Exam Tips appendix.
- **Providing over 1600 AP®-style exercises throughout the book.** Each chapter contains a mix of free-response and multiple-choice questions that are similar to those found on the AP® Statistics exam. At the start of each Chapter Wrap-Up, you will find a FRAPPY (Free Response AP® Problem, Yay!). Each FRAPPY gives you the chance to solve an AP®-style free-response problem based on the material in the chapter. After you finish, you can view and critique

two example solutions from the book's Student Site (highschool.bfwpub.com/updatedtps6e). Then you can score your own response using a rubric provided by your teacher.

- **Developing the Course Skills for AP® Statistics through frequent repetition.** See the inside back cover for more details.

Turn the page for a tour of the text. See how to use the book to realize success in the course and on the AP® Statistics exam.

READ THE TEXT and use the book's features to help you grasp the big ideas.

Read the **LEARNING TARGETS** at the beginning of each section. Focus on mastering these skills and concepts as you work through the chapter.

Scan the margins for the green notes, which represent the "voice of the teacher" giving helpful hints for being successful in the course. Many of these notes include important reminders from the AP® Statistics Course Framework.

It's important to learn the language of statistics. Take note of the green **DEFINITION** boxes that explain important vocabulary. Flip back to them to review key terms and their definitions, or turn to the Glossary/Glosario at the back of the book.

Read the **AP® EXAM TIPS.** They give advice on how to be successful on the AP® Statistics exam.

Look for the boxes with the green bands. Some explain how to make graphs or set up calculations, while others recap important concepts.

Watch for **CAUTION ICONS.** They alert you to common mistakes that students often make.

SECTION 3.1 Scatterplots and Correlation

LEARNING TARGETS *By the end of the section, you should be able to:*

- Distinguish between explanatory and response variables for quantitative data.
- Make a scatterplot to display the relationship between two quantitative variables.
- Describe the direction, form, and strength of a relationship displayed in a scatterplot and identify unusual features.
- Interpret the correlation.
- Understand the basic properties of correlation, including how the correlation is influenced by unusual points.
- Distinguish correlation from causation.

A one-variable data set is sometimes called *univariate data*. A data set that describes the relationship between two variables is sometimes called *bivariate data*.

Most statistical studies examine data on more than one variable for a group of individuals. Fortunately, analysis of relationships between two variables builds on the same tools we used to analyze one variable. The principles that guide our work also remain the same:

- Plot the data, then look for overall patterns and departures from those patterns.
- Add numerical summaries.
- When there's a regular overall pattern, use a simplified model to describe it.

Explanatory and Response Variables

In the "Candy grab" activity, the number of candies is the **response variable.** Hand span is the **explanatory variable** because we anticipate that knowing a student's hand span will help us predict the number of candies that student can grab.

DEFINITION Response variable, Explanatory variable

A **response variable** measures an outcome of a study. An **explanatory variable** may help predict or explain changes in a response variable.

You will often see explanatory variables called *independent variables* and response variables called *dependent variables*. Because the words *independent* and *dependent* have other meanings in statistics, we won't use them here.

It is easiest to identify explanatory and response variables when we initially specify the values of one variable to see how it affects another variable. For instance, to study the effect of alcohol on body temperature, researchers gave several different amounts of alcohol to mice. Then they measured the change in each mouse's body temperature 15 minutes later. In this case, amount of alcohol is the explanatory variable, and change in body temperature is the response variable. When we don't specify the values of either variable before collecting the data, there may or may not be a clear explanatory variable.

AP® EXAM TIP

When you are asked to *describe* the association shown in a scatterplot, you are expected to discuss the direction, form, and strength of the association, along with any unusual features, *in the context of the problem*. This means that you need to use both variable names in your description.

HOW TO DESCRIBE A SCATTERPLOT

To describe a scatterplot, make sure to address the following four characteristics in the context of the data:

- **Direction:** A scatterplot can show a positive association, negative association, or no association.
- **Form:** A scatterplot can show a linear form or a nonlinear form. The form is linear if the overall pattern follows a straight line. Otherwise, the form is nonlinear.
- **Strength:** A scatterplot can show a weak, moderate, or strong association. An association is strong if the points don't deviate much from the form identified. An association is weak if the points deviate quite a bit from the form identified.

Few relationships are linear for all values of the explanatory variable. **Don't make predictions using values of *x* that are much larger or much smaller than those that actually appear in your data.**

LEARN STATISTICS BY *DOING* STATISTICS

Every chapter begins with a hands-on **ACTIVITY** that introduces the content of the chapter. Many of these activities involve collecting data and drawing conclusions from the data. In other activities, you'll use dynamic applets to explore statistical concepts.

ACTIVITY **Candy grab**

In this activity, you will investigate if students with a larger hand span can grab more candy than students with a smaller hand span.[1]

1. Measure the span of your dominant hand to the nearest half-centimeter (cm). Hand span is the distance from the tip of the thumb to the tip of the pinkie finger on your fully stretched-out hand.
2. One student at a time, go to the front of the class and use your dominant hand to grab as many candies as possible from the container. You must grab the candies with your fingers pointing down (no scooping!) and hold the candies for 2 seconds before counting them. After counting, put the candy back into the container.
3. On the board, record your hand span and number of candies in a table with the following headings:

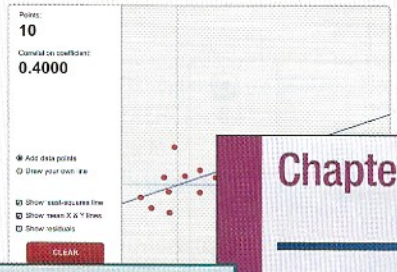
Hand span (cm)	Number of candies

4. While other students record their values on the board, copy the table onto a piece of paper and make a graph. Begin by constructing a set of coordinate axes. Label the horizontal axis the vertical axis "Number of candies." scale for each axis and plot each point e as accurately as you can on the graph.

ell you about the relationship between r of candies? Summarize your observa-o.

ACTIVITY **Investigating properties of the least-squares regression line**

In this activity, you will use the *Correlation and Regression* applet to explore some properties of the least-squares regression line.



1. Launch the applet at highschool.bfwpub.com/updatedtps6e.
2. Click on the graphing area to add 10 points in the lower-left corner so that the correlation is about $r = 0.40$. Also, check the boxes to show the "Least-Squares Line" and the "Mean X & Y"

Chapters 1, 3, 4, 11, and 12 conclude with a **CHAPTER PROJECT.** Three of the projects (Chapters 1, 3, and 11) provide an opportunity to think like a statistician by analyzing larger data sets with multiple variables of interest. The other two (Chapters 4 and 12) are longer-term projects that require you to engage in the statistical problem-solving process: Ask Questions, Collect Data, Analyze Data, Interpret Results.

Chapter 3 Project Investigating Relationships in Baseball

What is a better predictor of the number of wins for a baseball team, the number of runs scored by the team or the number of runs they allow the other team to score? What variables can we use to predict the number of runs a team scores? To predict the number of runs it allows the other team to score? In this project, you will use technology to help answer these questions by exploring a large set of data from Major League Baseball.

Part 1

1. Download the "MLB Team Data 2012–2016" Excel file from the book's website, along with the "Glossary for MLB Team Data file," which explains each of the variables included in the data set.[38] Import the data into th cal software package you prefer.
2. Create a scatterplot to investigate the relationship runs scored per game (R/G) and wins (W). The late the equation of the least-squares regression standard deviation of the residuals, and r^2. *Note:* the section for hitting statistics and W is in the sc pitching statistics.
5. Because the number of wins a team has is dependent on both how many runs they score and how many runs they allow, we can use a combination of both variables to predict the number of wins. Add a column in your data table for a new variable, run differential. Fill in the values using the formula R/G – RA/G.
6. Create a scatterplot to investigate the relationship between run differential and wins. Then calculate the equation of the least-squares regression line, the standard deviation of the residuals, and r^2.
7. Is run differential a better predictor than the variable you chose in Question 4? Explain your reasoning.

CHECK YOUR UNDERSTANDING

In Exercises 3 and 7, we asked you to make and describe a scatterplot for the hiker data shown in the table.

Body weight (lb)	120	187	109	103	131	165	158	116
Backpack weight (lb)	26	30	26	24	29	35	31	28

1. Calculate the equation of the least-squares regression line.
2. Make a residual plot for the linear model in Question 1.
3. What does the residual plot indicate about the appropriateness of the linear model? Explain your answer.

CHECK YOUR UNDERSTANDING questions appear throughout the section. They help clarify definitions, concepts, and procedures. Be sure to check your answers in the back of the book.

EXAMPLES: Model statistical problems and how to solve them

Read through each **EXAMPLE**, and then try out the concept yourself by working the **FOR PRACTICE, TRY** exercise in the Section Exercises.

Need extra help? Examples and exercises marked with the **PLAY ICON** are supported by short video clips prepared by experienced AP® Statistics teachers. The video guides you through each step in the example and solution and provides additional explanation when you need it.

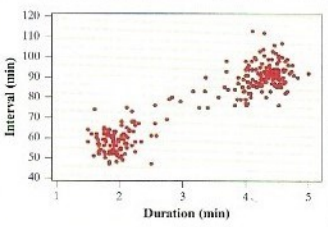
EXAMPLE Old Faithful and fertility
Describing a scatterplot

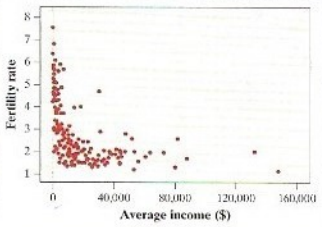

BigshotPOS/iStock/Getty Images

PROBLEM: Describe the relationship in each of the following contexts.

(a) The scatterplot on the left shows the relationship between the duration (in minutes) of an eruption and the interval of time until the next eruption (in minutes) of Old Faithful during a particular month.

(b) The scatterplot on the right shows the relationship between the average income (gross domestic product per person, in dollars) and fertility rate (number of children per woman) in 187 countries.[4]



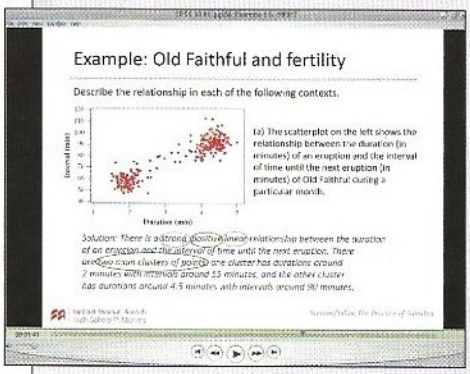


SOLUTION:

(a) There is a strong, positive linear relationship between the duration of an eruption and the interval of time until the next eruption. There are two main clusters of points: one cluster has durations around 2 minutes with in
has durations aro

(b) There is a modera
between average
is a country outsi
$30,000 and a f

Even with the clusters, the overall direction is still positive. In some cases, however, the points in a cluster



EXAMPLE Caffeine and pulse rates
How random assignment works

PROBLEM: A total of 20 students have agreed to participate in an experiment comparing the effects of caffeinated cola and caffeine-free cola on pulse rates. Describe how you would randomly assign 10 students to each of the two treatments:

(a) Using 20 identical slips of paper

(b) Using technology

(c) Using Table D


Zeke Sikelianos

SOLUTION:

(a) On 10 slips of paper, write the letter "A"; on the remaining 10 slips, write the letter "B." Shuffle the slips of paper and hand out one slip of paper to each volunteer. Students who get an "A" slip receive the cola with caffeine and students who get a "B" slip receive the cola without caffeine.

When describing a method of random assignment, don't stop after creating the groups. Make sure to identify which group gets which treatment.

(b) Label each student with a different integer from 1 to 20. Then randomly generate 10 different integers from 1 to 20. The students with these labels receive the cola with caffeine. The remaining 10 students receive the cola without caffeine.

When using a random number generator or a table of random digits to assign treatments, make sure to account for the possibility of repeated numbers when describing your method.

(c) Label each student with a different integer from 01 to 20. Go to a line of Table D and read two-digit groups moving from left to right. The first 10 different labels between 01 and 20 identify the 10 students who receive cola with caffeine. The remaining 10 students receive the caffeine-free cola. Ignore groups of digits from 21 to 00.

FOR PRACTICE, TRY EXERCISE 63

The **SOLUTION** is presented in a special font and models the style, steps, and language that you should use to earn full credit on the AP® Statistics exam.

THE VOICE OF THE TEACHER. Study the worked examples and pay special attention to the carefully placed **"Teacher Talk" comment boxes** that guide you step by step through the solution. These comments offer lots of good advice—as if your teacher is working directly with you to solve a problem.

The blue page number icon next to an exercise points you back to the page on which the model example appears.

63. **Layoffs and "survivor guilt"** Workers who survive a pg 251 layoff of other employees at their location may suffer from "survivor guilt." A study of survivor guilt and its effects used as subjects 120 students who were offered an opportunity to earn extra course credit by doing proofreading. Each subject worked in the same cubicle as another student, who was an accomplice of the experimenters. At a break midway through the work,

EXERCISES: Practice makes perfect!

Start by reading the **SECTION SUMMARY** to be sure that you understand the big ideas and key concepts.

Section 3.1 Summary

- A **scatterplot** displays the relationship between two quantitative variables measured on the same individuals. Mark values of one variable on the horizontal axis (x axis) and values of the other variable on the vertical axis (y axis). Plot each individual's data as a point on the graph.
- If we think that a variable x may help predict, explain, or even cause changes in another variable y, we call x an **explanatory variable** and y a **response variable.** Always plot the explanatory variable on the x axis of a scatterplot. Plot the response variable on the y axis.
- When describing a scatterplot, look for an overall pattern (direction, form, strength) and departures from the pattern (unusual features) and always answer in cor
  - Direction
    variable

Practice! Work the **EXERCISES** assigned by your teacher. Compare your answers to those in the Solutions appendix at the back of the book. Short solutions to the exercises numbered in red are found in the appendix.

Most of the **exercises are paired**, meaning that odd- and even-numbered exercises test the same skill or concept. If you answer an assigned exercise incorrectly, try to figure out your mistake. Then see if you can solve the paired exercise.

Section 3.1 Exercises

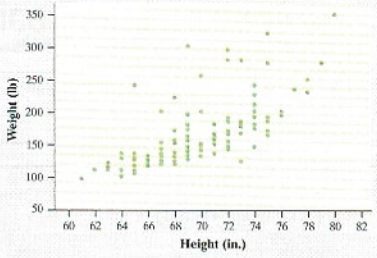
1. **Coral reefs and cell phones** Identify the explanatory pg 154 variable and the response variable for the following relationships, if possible. Explain your reasoning.
   (a) The weight gain of corals in aquariums where the water temperature is controlled at different levels
   (b) The number of text messages sent and the number of phone calls made in a sample of 100 students
2. **Teenagers and corn yield** Identify the explanatory variable and the response variable for the following relationships, if possible. Explain your reasoning.
   (a) The height and arm span of a sample of 50 teenagers
   (b) The yield of corn in bushels per acre and the amount of rain in the growing season
3. **Heavy backpacks** Ninth-grade students at the Webb pg 155 Schools go on a backpacking trip each fall. Students are divided into hiking groups of size 8 by selecting names from a hat. Before leaving, students and their backpacks are weighed. The data here are from one hiking group. Make a scatterplot by hand that shows how backpack weight relates to body weight.

Body weight (lb)	120	187	109	103	131	165	158	116
Backpack weight (lb)	26	30	26	24	29	35	31	28

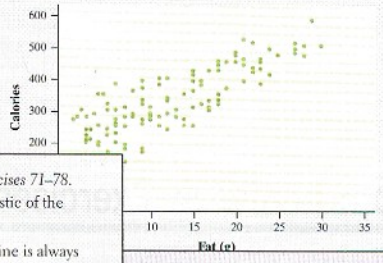
4. **Putting success** How well do professional golfers putt from various distances to the hole? The data show various distances to the hole (in feet) and the percent of putts mad
   Make a scatt
   of putts mad

Distance (ft)	

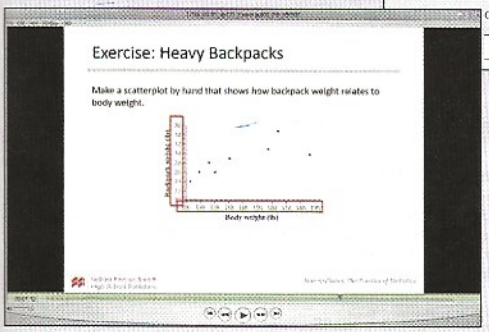
Track and Field team.[10] Describe the relationship between height and weight for these athletes.



6. **Starbucks** The scatterplot shows the relationship between the amount of fat (in grams) and number of calories in products sold at Starbucks.[11] Describe the relationship between fat and calories for these products.



Look for **ICONS** that appear next to selected **EXERCISES**. They will guide you to
- the Example that models the exercise.
- videos that provide step-by-step instructions for solving the exercise.



Multiple Choice: *Select the best answer for Exercises 71–78.*

71. Which of the following is *not* a characteristic of the least-squares regression line?
   (a) The slope of the least-squares regression line is always between −1 and 1.
   (b) The least-squares regression line always goes through the point $(\bar{x}, \bar{y})$.
   (c) The least-squares regression line minimizes the sum of squared residuals.
   (d) The slope of the least-squares regression line w always have the same sign as the correlation.
   (e) The least-squares regression line is not resistant outliers.

Recycle and Review

79. **Fuel economy** (2.2) In its recent *Fuel Economy Guide*, the Environmental Protection Agency (EPA) gives data on 1152 vehicles. There are a number of outliers, mainly vehicles with very poor gas mileage or hybrids with very good gas mileage. If we ignore the outliers, however, the combined city and highway gas mileage of the other 1120 or so vehicles is approximately Normal with mean 18.7 miles per gallon (mpg) and standard deviation 4.3 mpg.
   (a) The Chevrolet Malibu with a four-cylinder engine has a combined gas mileage of 25 mpg. What percent of the 1120 vehicles have worse gas mileage than the Malibu?

Various types of problems in the Section Exercises let you practice solving many different types of questions, including AP®-style **multiple-choice** and **free-response**. The **Recycle and Review** exercises refer back to concepts and skills learned in an earlier section, noted in purple after the problem title.

REVIEW and PRACTICE for quizzes and tests

Study the **CHAPTER REVIEW** to be sure that you understand the key concepts in each section.

Chapter 3 Wrap-Up

Chapter 3 Review

Section 3.1: Scatterplots and Correlation

In this section, you learned how to explore the relationship between two quantitative variables. As with distributions of a single variable, the first step is always to make a graph. A scatterplot is the appropriate type of graph to investigate relationships between two quantitative variables. To describe a scatterplot, be sure to discuss four characteristics: direction, form, strength, and unusual features. The direction of a

Use the **WHAT DID YOU LEARN?** table that directs you to examples and exercises to verify your mastery of each **LEARNING TARGET**.

What Did You Learn?

Learning Target	Section	Related Example on Page(s)	Relevant Chapter Review Exercise(s)
Distinguish between explanatory and response variables for quantitative data.	3.1	154	R3.4
Make a scatterplot to display the relationship between two quantitative variables.	3.1	155	R3.4

SUMMARY TABLES in Chapters 8–12 review important details of each inference procedure, including conditions and formulas.

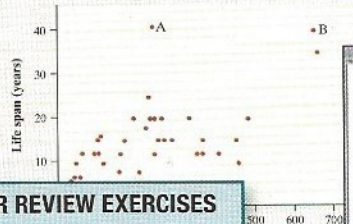
Comparing confidence intervals for proportions

	Confidence interval for p	Confidence interval for $p_1 - p_2$
Name (TI-83/84)	One-sample z interval for p (1-PropZInt)	Two-sample z interval for $p_1 - p_2$ (2-PropZInt)
Conditions	• **Random:** The data come from a random sample from the population of interest. ◦ **10%:** When sampling without replacement, $n < 0.10N$. • **Large Counts:** Both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10. That is, the number of successes and the number of failures in the sample are both at least 10.	• **Random:** The data come from two independent random samples or from two groups in a randomized experiment. ◦ **10%:** When sampling without replacement, $n_1 < 0.10N_1$ and $n_2 < 0.10N_2$. • **Large Counts:** The counts of "successes" and "failures" in each sample or group—$n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, $n_2(1-\hat{p}_2)$—are all at least 10.
Formula	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

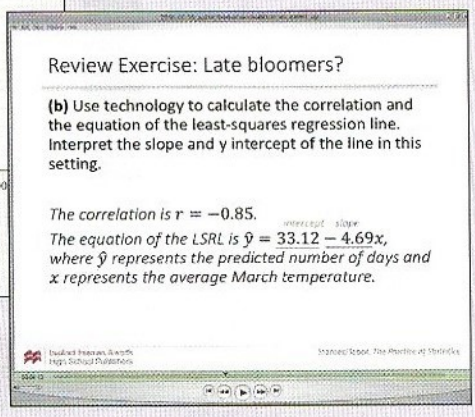
Chapter 3 Review Exercises

These exercises are designed to help you review the important ideas and methods of the chapter.

R3.1 **Born to be old?** Is there a relationship between the gestational period (time from conception to birth) of an animal and its average life span? The figure shows a scatterplot of the gestational period and average life span for 43 species of animals.[53]



Tackle the **CHAPTER REVIEW EXERCISES** for practice in solving problems that test concepts from throughout the chapter. Need more help or just want additional insights before you take the practice test? Watch the **Chapter Review Exercise Videos**.



and the AP® STATISTICS EXAM

Chapter 3 AP® Statistics Practice Test

Section I: Multiple Choice *Select the best answer for each question.*

T3.1 A school guidance counselor examines how many extracurricular activities students participate in and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to 0." Which of the following is the most appropriate conclusion?

(a) Students involved in many extracurricular activities tend to be students with poor grades.

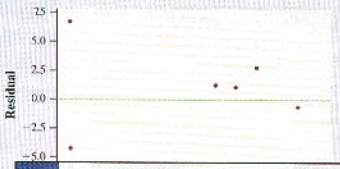
(b) Students with good grades tend to be students who are not involved in many extracurricular activities.

(c) Students involved in many extracurricular activities are just as likely to get good grades as bad grades.

(d) Students with good grades tend to be students who are involved in many extracurricular activities.

(e) No conclusion should be made based on the correlation without looking at a scatterplot of the data.

Questions T3.3–T3.5 refer to the following setting. Scientists examined the activity level of 7 fish at different temperatures. Fish activity was rated on a scale of 0 (no activity) to 100 (maximal activity). The temperature was measured in degrees Celsius. A computer regression printout and a residual plot are provided. Notice that the horizontal axis on the residual plot is labeled "Fitted value," which means the same thing as "predicted value."



Each chapter concludes with an **AP® STATISTICS PRACTICE TEST**. This test includes about 10 AP®-style multiple-choice questions and 3 free-response questions.

Cumulative AP® Practice Test 1

Section I: Multiple Choice *Choose the best answer for Questions AP1.1–AP1.14.*

AP1.1 You look at real estate ads for houses in Sarasota, Florida. Many houses have prices from $200,000 to $400,000. The few houses on the water, however, have prices up to $15 million. Which of the following statements best describes the distribution of home prices in Sarasota?

(a) The distribution is most likely skewed to the left, and the mean is greater than the median.

(b) The distribution is most likely skewed to the left, and the mean is less than the median.

(c) The distribution is roughly symmetric with a few high outliers, and the mean is approximately equal to the median.

(d) The distribution is most likely skewed to the right, and the mean is greater than the median.

(e) The distribution is most likely skewed to the right, and the mean is less than the median.

AP1.2 A child is 40 inches tall, which places her at the 90th percentile of all children of similar age. The heights for children of this age form an approximately Normal distribution with a mean of 38 inches. Based on this information, what is the standard deviation of the heights of all children of this age?

(a) 0.20 inch

(b) 0.31 inch

(c) 0.65 inch

(d) 1.21 inches

(e) 1.56 inches

Four **CUMULATIVE AP® PRACTICE TESTS** simulate the real exam. They are placed after Chapters 4, 7, 11, and 12. The tests expand in length and content coverage as you work through the book. The last test models a full AP® Statistics exam.

FRAPPY! FREE RESPONSE AP® PROBLEM, YAY!

The following problem is modeled after actual AP® Statistics exam free response questions. Your task is to generate a complete, concise response in 15 minutes.

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

Two statistics students went to a flower shop and randomly selected 12 carnations. When they got home, the students prepared 12 identical vases with exactly the same amount of water in each vase. They put one tablespoon of sugar in 3 vases, two tablespoons of sugar in 3 vases, and three tablespoons of sugar in 3 vases. In the remaining 3 vases, they put no sugar. After the vases were prepared, the students randomly assigned 1 carnation to each vase and observed how many hours each flower continued to look fresh. A scatterplot of the data is shown below.

(a) Briefly describe the association shown in the scatterplot.

(b) The equation of the least-squares regression line for these data is $\hat{y} = 180.8 + 15.8x$. Interpret the slope of the line in the context of the study.

(c) Calculate and interpret the residual for the flower that had 2 tablespoons of sugar and looked fresh for 204 hours.

(d) Suppose that another group of students conducted a similar experiment using 12 flowers, but included different varieties in addition to carnations. Would you expect the value of r^2 for the second group's data to be greater than, less than, or about the same as the value of r^2 for the first group's data? Explain.

After you finish, you can view two example solutions on the book's website (highschool.bfwpub.com/updatedtps6e). Determine whether you think each solution is "complete," "substantial," "developing," or "minimal." If the solution is not complete, what improvements would you suggest to the student who wrote it? Finally, your teacher will provide you with a scoring rubric. Score your response and note what, if anything, you would do differently

Learn how to answer free response questions successfully by working the **FRAPPY!**—the Free Response AP® Problem, Yay!—that begins the Chapter Wrap-Up in every chapter.

Use TECHNOLOGY to discover and analyze

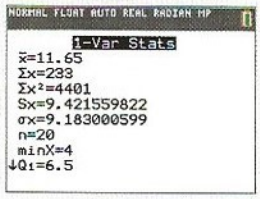
3. Technology Corner COMPUTING NUMERICAL SUMMARIES

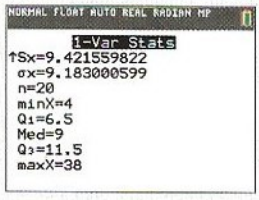
TI-Nspire and other technology instructions are on the book's website at highschool.bfwpub.com/updatedtps6e.

Let's find numerical summaries for the boys' shoes data from the example on page 64. We'll start by showing you how to compute summary statistics on the TI-83/84 and then look at output from computer software.

I. **One-variable statistics on the TI-83/84**

1. Enter the data in list L1.
2. Find the summary statistics for the shoe data.
   - Press STAT (CALC); choose 1-VarStats.
     OS 2.55 or later: In the dialog box, press 2nd 1 (L1) and ENTER to specify L1 as the List. Leave FreqList blank. Arrow down to Calculate and press ENTER.
     Older OS: Press 2nd 1 (L1) and ENTER.
   - Press ▼ to see the rest of the one-variable statistics.





II. **Output from statistical software** We used Minitab statistical software to calculate descriptive statistics for the boys' shoes data. Minitab allows you to choose which numerical summaries are included in the output.

Descriptive Statistics: Shoes

Variable	N	Mean	StDev	Minimum	Q_1	Median	Q_3	Maximum
Shoes	20	11.65	9.42	4.00	6.25	9.00	11.75	38.00

Note: The TI-83/84 gives the first and third quartiles of the boys' shoes distribution as $Q_1 = 6.5$ and $Q_3 = 11.5$. Minitab reports that $Q_1 = 6.25$ and $Q_3 = 11.75$. What happened? Minitab and some other software use slightly different rules for locating quartiles. Results from the various rules are usually close to each other. Be aware of possible differences when calculating quartiles as they may affect more than just the *IQR*.

Use technology as a tool for discovery and analysis. **TECHNOLOGY CORNERS** give step-by-step instructions for using the TI-83/84 calculator. Instructions for the TI-Nspire and other calculators are on the book's Student Site (highschool.bfwpub.com/updatedtps6e) and in the e-Book platform.

Technology Corner videos are also available to walk you through the key strokes needed to perform each analysis.

Although the Technology Corners focus on the TI-83/84 graphing calculator, output from multiple programs—including Minitab and JMP—is used in the book's Examples and Exercises to help you become familiar with reading and interpreting many different kinds of statistical summaries.

3.2 Technology Corners

TI-Nspire and other technology instructions are on the book's website at highschool.bfwpub.com/updatedtps6e.

Find the Technology Corners easily by consulting the summary table at the end of each section or the complete table at the back of the book.

Read, practice, access the resources, and do homework assignments online with the new **Online Homework and e-Book Platform** that may be purchased to enhance your learning experience.

OVERVIEW: What Is Statistics?

Does listening to music while studying help or hinder learning? If an athlete fails a drug test, how sure can we be that she took a banned substance? Does having a pet help people live longer? How well do SAT scores predict college success? Do most people recycle? Which of two diets will help obese children lose more weight and keep it off? Can a new drug help people quit smoking? How strong is the evidence for global warming?

These are just a few of the questions that statistics can help answer. But what is statistics? And why should you study it?

Statistics Is the Science of Learning from Data

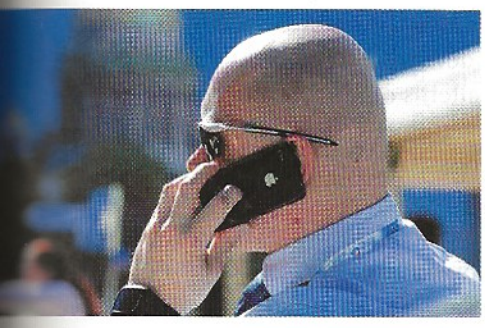
istockphoto

Data are usually numbers, but they are not "just numbers." *Data are numbers with a context.* The number 10.5, for example, carries no information by itself. But if we hear that a family friend's new baby weighed 10.5 pounds at birth, we congratulate her on the healthy size of the child. The context engages our knowledge about the world and allows us to make judgments. We know that a baby weighing 10.5 pounds is quite large, and that a human baby is unlikely to weigh 10.5 ounces or 10.5 kilograms. The context makes the number meaningful.

In your lifetime, you will be bombarded with data and statistical information. Poll results, television ratings, music sales, gas prices, unemployment rates, medical study outcomes, and standardized test scores are discussed daily in the media. Using data effectively is a large and growing part of most professions. A solid understanding of statistics will enable you to make sound, data-based predictions, decisions, and conclusions in your career and everyday life.

Data Beat Personal Experiences

It is tempting to base conclusions on your own experiences or the experiences of those you know. But our experiences may not be typical. In fact, the incidents that stick in our memory are often the unusual ones.



Bloomberg via Getty Images

Do Cell Phones Cause Brain Cancer?

Italian businessman Innocente Marcolini developed a brain tumor at age 60. He also talked on a cellular phone up to 6 hours per day for 12 years as part of his job. Mr. Marcolini's physician suggested that the brain tumor may have been caused by cell-phone use. So Mr. Marcolini decided to file suit in the Italian court system. A court ruled in his favor in October 2012.

Several statistical studies have investigated the link between cell-phone use and brain cancer. One of the largest was conducted by the Danish Cancer Society.

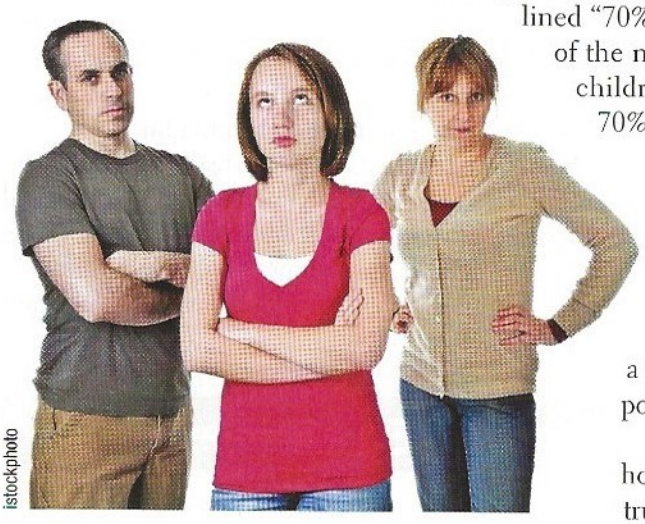
Over 350,000 residents of Denmark were included in the study. Researchers compared the brain-cancer rate for the cell-phone users with the rate in the general population. The result: no statistical difference in brain-cancer rates.[1] In fact, most studies have produced similar conclusions. In spite of the evidence, many people (like Mr. Marcolini) are still convinced that cell phones can cause brain cancer.

In the public's mind, the compelling story wins every time. A statistically literate person knows better. ***Data are more reliable than personal experiences because they systematically describe an overall picture, rather than focus on a few incidents.***

Where the Data Come from Matters

Are You Kidding Me?

The famous advice columnist Ann Landers once asked her readers, "If you had it to do over again, would you have children?" A few weeks later, her column was headlined "70% OF PARENTS SAY KIDS NOT WORTH IT." Indeed, 70% of the nearly 10,000 parents who wrote in said they would not have children if they could make the choice again. Do you believe that 70% of all parents regret having children?



istockphoto

You shouldn't. The people who took the trouble to write to Ann Landers are not representative of all parents. Their letters showed that many of them were angry with their children. All we know from these data is that there are some unhappy parents out there. A statistically designed poll, unlike Ann Landers's appeal, targets specific people chosen in a way that gives all parents the same chance to be asked. Such a poll showed that 91% of parents *would* have children again.

Where data come from matters a lot. If you are careless about how you get your data, you may announce 70% "No" when the truth is close to 90% "Yes."

Who Talks More—Women or Men?

According to Louann Brizendine, author of *The Female Brain*, women say nearly 3 times as many words per day as men. Skeptical researchers devised a study to test this claim. They used electronic devices to record the talking patterns of 396 university students from Texas, Arizona, and Mexico. The device was programmed to record 30 seconds of sound every 12.5 minutes without the carrier's knowledge. What were the results?

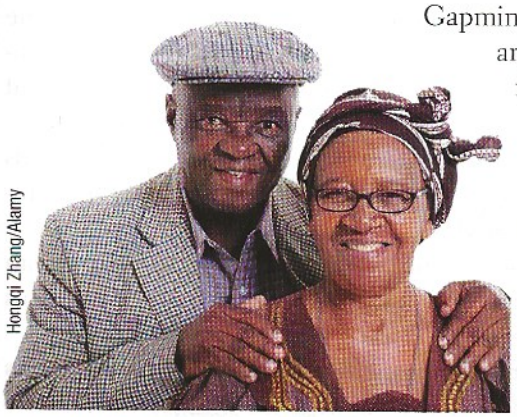
According to a published report of the study in *Scientific American*, "Men showed a slightly wider variability in words uttered. . . . But in the end, the sexes came out just about even in the daily averages: women at 16,215 words and men at 15,669."[2] When asked where she got her figures, Brizendine admitted that she used unreliable sources.[3]

The most important information about any statistical study is how the data were produced. Only carefully designed studies produce results that can be trusted.

Always Plot Your Data

Yogi Berra, a famous New York Yankees baseball player known for his unusual quotes, had this to say: "You can observe a lot just by watching." That's a motto for learning from data. *A carefully chosen graph helps us describe patterns in data and identify important departures from those patterns.*

Do People Live Longer in Wealthier Countries?

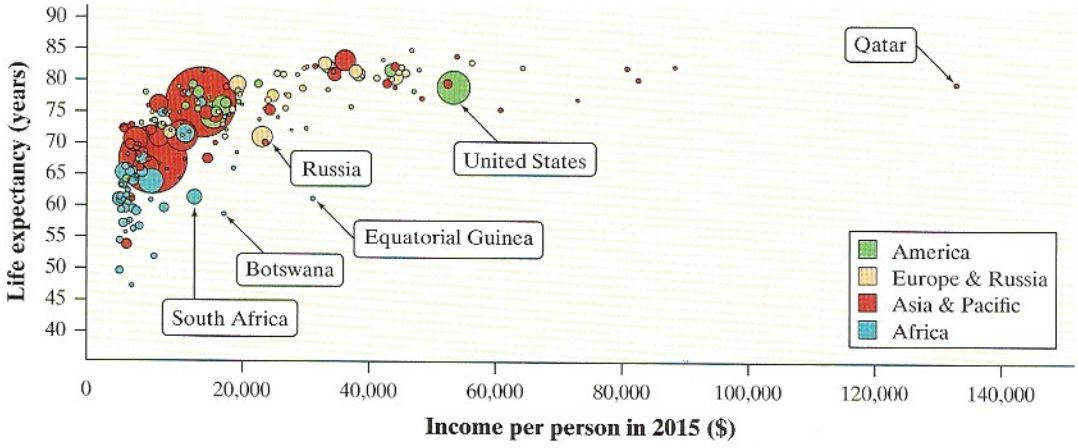


Hongqi Zhang/Alamy

The Gapminder website, www.gapminder.org, provides loads of data on the health and well-being of the world's inhabitants. The graph below displays some data from Gapminder.[4] The individual points represent all the world's nations for which data are available. Each point shows the income per person and life expectancy for one country, along with the region (color of point) and population (size of point).

We expect people in richer countries to live longer. The overall pattern of the graph does show this, but the relationship has an interesting shape. Life expectancy rises very quickly as personal income increases and then levels off. People in very rich countries like the United States live no longer than people in poorer but not extremely poor nations. In some less wealthy countries, people live longer than in the United States. Several other nations stand out in the graph. What's special about each of these countries?

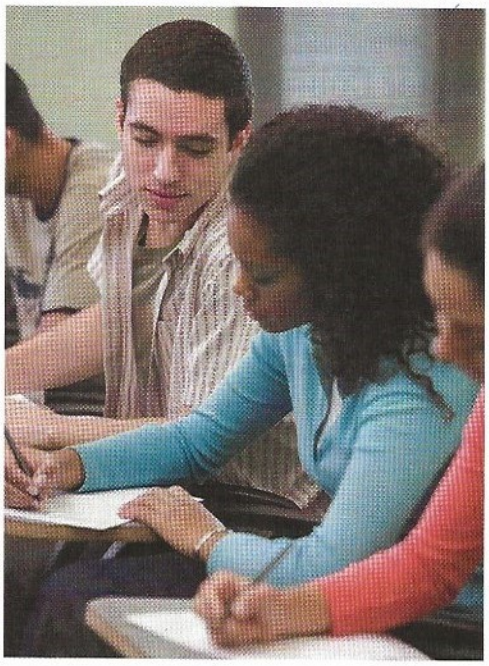
Graph of the life expectancy of people in many nations against each nation's income per person in 2015.



Variation Is Everywhere

Individuals vary. Repeated measurements on the same individual vary. Chance outcomes—like spins of a roulette wheel or tosses of a coin—vary. Almost everything varies over time. Statistics provides tools for understanding variation.

Have Most Students Cheated on a Test?



Commercial Eye/Getty Images

Researchers from the Josephson Institute were determined to find out. So they surveyed about 23,000 students from 100 randomly selected schools (both public and private) nationwide. The question was: "How many times have you cheated during a test at school in the past year?" Fifty-one percent said they had cheated at least once.[5]

If the researchers had asked the same question of *all* high school students, would exactly 51% have answered "Yes"? Probably not. If the Josephson Institute had selected a different sample of about 23,000 students to respond to the survey, they would probably have gotten a different estimate. *Variation is everywhere!*

Fortunately, statistics provides a description of how the sample results will vary in relation to the actual population percent. Based on the sampling method that this study used, we can say that the estimate of 51% is very likely to be within 1% of the true population value. That is, we can be quite confident that between 50% and 52% of *all* high school students would say that they have cheated on a test.

Because variation is everywhere, conclusions are uncertain. Statistics gives us the tools to quantify our uncertainty, allowing for valid, data-based predictions, decisions, and conclusions.

UNIT 1
Exploring One-Variable Data

Chapter 1

Data Analysis

AfriPics.com/Alamy Stock Photo

INTRODUCTION Statistics: The Science and Art of Data

LEARNING TARGETS *By the end of the section, you should be able to:*

- Identify the individuals and variables in a set of data.
- Classify variables as categorical or quantitative.

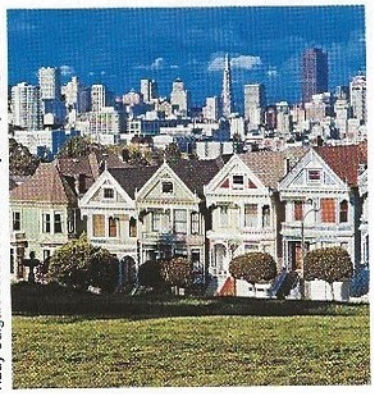
We live in a world of *data*. Every day, the media report poll results, outcomes of medical studies, and analyses of data on everything from stock prices to standardized test scores to global warming. The data are trying to tell us a story. To understand what the data are saying, you need to learn more about **statistics.**

DEFINITION Statistics

Statistics is the science and art of collecting, analyzing, and drawing conclusions from data.

A solid understanding of statistics will help you make good decisions based on data in your daily life.

Organizing Data



Rudy Sulgan/Corbis Documentary/Getty Images

Every year, the U.S. Census Bureau collects data from over 3 million households as part of the American Community Survey (ACS). The table displays some data from the ACS in a recent year.

Household	Region	Number of people	Time in dwelling (years)	Response mode	Household income	Internet access?
425	Midwest	5	2–4	Internet	52,000	Yes
936459	West	4	2–4	Mail	40,500	Yes
50055	Northeast	2	10–19	Internet	481,000	Yes
592934	West	4	2–4	Phone	230,800	No
545854	South	9	2–4	Phone	33,800	Yes
809928	South	2	30+	Internet	59,500	Yes
110157	Midwest	1	5–9	Internet	80,000	Yes
999347	South	1	<1	Mail	8,400	No

Most data tables follow this format—each row describes an **individual** and each column holds the values of a **variable.**

Sometimes the individuals in a data set are called *cases* or *observational units*.

DEFINITION Individual, Variable

An **individual** is an object described in a set of data. Individuals can be people, animals, or things.

A **variable** is an attribute that can take different values for different individuals.

For the American Community Survey data set, the *individuals* are households. The *variables* recorded for each household are region, number of people, time in current dwelling, survey response mode, household income, and whether the dwelling has Internet access. Region, time in dwelling, response mode, and Internet access status are **categorical variables**. Number of people and household income are **quantitative variables**.

Note that household is *not* a variable. The numbers in the household column of the data table are just labels for the individuals in this data set. Be sure to look for a column of labels—names, numbers, or other identifiers—in any data table you encounter.

AP® EXAM TIP

If you learn to distinguish categorical from quantitative variables now, it will pay big rewards later. You will be expected to analyze categorical and quantitative variables correctly on the AP® exam.

DEFINITION Categorical variable, Quantitative variable

A **categorical variable** assigns labels that place each individual into a particular group, called a category.

A **quantitative variable** takes number values that are quantities—counts or measurements.

Not every variable that takes number values is quantitative. Zip code is one example. Although zip codes are numbers, they are neither counts of anything, nor measurements of anything. They are simply labels for a regional location, making zip code a categorical variable. Some variables—such as gender, race, and occupation—are categorical by nature. Time in dwelling from the ACS data set is also a categorical variable because the values are recorded as intervals of time, such as 2–4 years. If time in dwelling had been recorded to the nearest year for each household, this variable would be quantitative.

To make life simpler, we sometimes refer to *categorical data* or *quantitative data* instead of identifying the variable as categorical or quantitative.

EXAMPLE Census At School
Individuals and Variables

PROBLEM: Census At School is an international project that collects data about primary and secondary school students using surveys. Hundreds of thousands of students from Australia, Canada, Ireland, Japan, New Zealand, South Africa, South Korea, the United Kingdom, and the United States have taken part in the project. Data from the surveys are available online. We used the site's "Random Data Selector" to choose 10 Canadian students who completed the survey in a recent year. The table displays the data.

Garry Black/Alamy

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circumference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

(a) Identify the individuals in this data set.

(b) What are the variables? Classify each as categorical or quantitative.

SOLUTION:

(a) 10 randomly selected Canadian students who participated in the Census At School survey.

(b) Categorical: Province, gender, handedness, preferred communication method

Quantitative: Number of languages spoken, height (cm), wrist circumference (mm)

We'll see in Chapter 4 why choosing at random, as we did in this example, is a good idea.

There is at least one suspicious value in the data table. We doubt that the girl who is 166 cm tall really has a wrist circumference of 65 mm (about 2.6 inches). Always look to be sure the values make sense!

FOR PRACTICE, TRY EXERCISE 1

There are two types of quantitative variables: *discrete* and *continuous*. Most **discrete variables** result from counting something, like the number of languages spoken in the preceding example. **Continuous variables** typically result from measuring something, like height or wrist circumference. Be sure to report the units of measurement (like centimeters for height and millimeters for wrist circumference) for a continuous variable.

DEFINITION Discrete variable, Continuous variable

A quantitative variable that takes a fixed set of possible values with gaps between them is a **discrete variable**.

A quantitative variable that can take any value in an interval on the number line is a **continuous variable**.

The proper method of data analysis depends on whether a variable is categorical or quantitative. For that reason, it is important to distinguish these two types of variables. The type of data determines what kinds of graphs and which numerical summaries are appropriate.

ANALYZING DATA A variable generally takes values that vary (hence the name *variable*!). Categorical variables sometimes have similar counts in each category and sometimes don't. For instance, we might have expected similar numbers of

males and females in the Census At School data set. But we aren't surprised to see that most students are right-handed. Quantitative variables may take values that are very close together or values that are quite spread out. We call the pattern of variation of a variable its **distribution**.

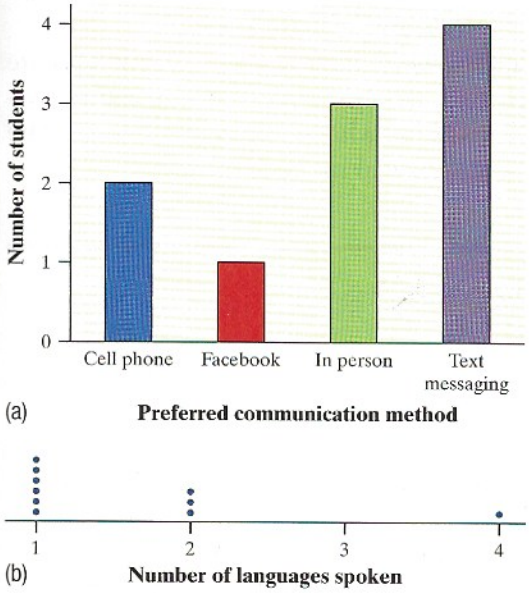


FIGURE 1.1 (a) Bar graph showing the distribution of preferred communication method for the sample of 10 Canadian students. (b) Dotplot showing the distribution of number of languages spoken by these students.

DEFINITION Distribution

The **distribution** of a variable tells us what values the variable takes and how often it takes those values.

Let's return to the data for the sample of 10 Canadian students from the preceding example. Figure 1.1(a) shows the distribution of preferred communication method for these students in a *bar graph*. We can see how many students chose each method from the heights of the bars: cell phone (2), Facebook (1), in person (3), text messaging (4). Figure 1.1(b) shows the distribution of number of languages spoken in a *dotplot*. We can see that 6 students speak one language, 3 students speak two languages, and 1 student speaks four languages.

Section 1.1 begins by looking at how to describe the distribution of a single categorical variable and then examines relationships between categorical variables. Sections 1.2 and 1.3 and all of Chapter 2 focus on describing the distribution of a quantitative variable. Chapter 3 investigates relationships between two quantitative variables. In each case, we begin with graphical displays, then add numerical summaries for a more complete description.

HOW TO ANALYZE DATA

- Begin by examining each variable by itself. Then move on to study relationships among the variables.
- Start with a graph or graphs. Then add numerical summaries.

This process of exploratory data analysis is known as *descriptive statistics*.

CHECK YOUR UNDERSTANDING

Jake is a car buff who wants to find out more about the vehicles that his classmates drive. He gets permission to go to the student parking lot and record some data. Later, he does some Internet research on each model of car he found. Finally, Jake makes a spreadsheet that includes each car's license plate, model, number of cylinders, color, highway gas mileage, weight, and whether it has a navigation system.

1. Identify the individuals in Jake's study.
2. What are the variables? Classify each as categorical or quantitative.
3. Identify each quantitative variable as discrete or continuous.

From Data Analysis to Inference

Sometimes we're interested in drawing conclusions that go beyond the data at hand. That's the idea of *inferential statistics*. In the "Census At School" example, 9 of the 10 randomly selected Canadian students are right-handed. That's 90% of the *sample*. Can we conclude that exactly 90% of the *population* of Canadian students who participated in Census At School are right-handed? No.

If another random sample of 10 students were selected, the percent who are right-handed might not be exactly 90%. Can we at least say that the actual population value is "close" to 90%? That depends on what we mean by "close." The following activity gives you an idea of how statistical inference works.

ACTIVITY Hiring discrimination—it just won't fly!

An airline has just finished training 25 pilots—15 male and 10 female—to become captains. Unfortunately, only eight captain positions are available right now. Airline managers announce that they will use a lottery to determine which pilots will fill the available positions. The names of all 25 pilots will be written on identical slips of paper. The slips will be placed in a hat, mixed thoroughly, and drawn out one at a time until all 8 captains have been identified.

Chojja/Getty Images

A day later, managers announce the results of the lottery. Of the 8 captains chosen, 5 are female and 3 are male. Some of the male pilots who weren't selected suspect that the lottery was not carried out fairly. One of these pilots asks your statistics class for advice about whether to file a grievance with the pilots' union.

The key question in this possible discrimination case seems to be: *Is it plausible (believable) that these results happened just by chance?* To find out, you and your classmates will *simulate* the lottery process that airline managers said they used.

1. Your teacher will give you a bag with 25 beads (15 of one color and 10 of another) or 25 slips of paper (15 labeled "M" and 10 labeled "F") to represent the 25 pilots. Mix the beads/slips thoroughly. Without looking, remove 8 beads/slips from the bag. Count the number of female pilots selected. Then return the beads/slips to the bag.
2. Your teacher will draw and label a number line for a class *dotplot*. On the graph, plot the number of females you got in Step 1.
3. Repeat Steps 1 and 2 if needed to get a total of at least 40 simulated lottery results for your class.
4. Discuss the results with your classmates. Does it seem plausible that airline managers conducted a fair lottery? What advice would you give the male pilot who contacted you?

Our ability to do inference is determined by how the data are produced. Chapter 4 discusses the two main methods of data production—sampling and

experiments—and the types of conclusions that can be drawn from each. As the activity illustrates, the logic of inference rests on asking, "What are the chances?" *Probability*, the study of chance behavior, is the topic of Chapters 5–7. We'll introduce the most common inference techniques in Chapters 8–12.

Introduction Summary

- **Statistics** is the science and art of collecting, analyzing, and drawing conclusions from data.
- A data set contains information about a number of **individuals.** Individuals may be people, animals, or things. For each individual, the data give values for one or more **variables.** A variable describes some characteristic of an individual, such as a person's height, gender, or salary.
- A **categorical variable** assigns a label that places each individual in one of several groups, such as male or female. A **quantitative variable** has numerical values that count or measure some characteristic of each individual, such as number of siblings or height in meters.
- There are two types of quantitative variables: discrete and continuous. A **discrete variable** has a fixed set of possible numeric values with gaps between them. A **continuous variable** can take any value in an interval on the number line. Discrete variables usually result from counting something; continuous variables usually result from measuring something.
- The **distribution** of a variable describes what values the variable takes and how often it takes them.

Introduction Exercises

The solutions to all exercises numbered in red may be found in the Solutions Appendix, starting on page S-1.

1. **A class survey** Here is a small part of the data set that describes the students in an AP® Statistics class. The data come from anonymous responses to a questionnaire filled out on the first day of class.

pg 3

Gender	Grade level	GPA	Children in family	Homework last night (min)	Android or iPhone?
F	9	2.3	3	0–14	iPhone
M	11	3.8	6	15–29	Android
M	10	3.1	2	15–29	Android
F	10	4.0	1	45–59	iPhone
F	10	3.4	4	0–14	iPhone
F	10	3.0	3	30–44	Android
M	9	3.9	2	15–29	iPhone
M	12	3.5	2	0–14	iPhone

(a) Identify the individuals in this data set.

(b) What are the variables? Classify each as categorical or quantitative.

2. **Coaster craze** Many people like to ride roller coasters. Amusement parks try to increase attendance by building exciting new coasters. The following table displays data on several roller coasters that were opened in a recent year.[1]

Roller coaster	Type	Height (ft)	Design	Speed (mph)	Duration (sec)
Wildfire	Wood	187.0	Sit down	70.2	120
Skyline	Steel	131.3	Inverted	50.0	90
Goliath	Wood	165.0	Sit down	72.0	105
Helix	Steel	134.5	Sit down	62.1	130
Banshee	Steel	167.0	Inverted	68.0	160
Black Hole	Steel	22.7	Sit down	25.5	75

(a) Identify the individuals in this data set.

(b) What are the variables? Classify each as categorical or quantitative.

3. **Hit movies** According to the Internet Movie Database, *Avatar* is tops based on box-office receipts worldwide as of January 2017. The following table displays data on several popular movies. Identify the individuals

and variables in this data set. Classify each variable as categorical or quantitative.

Movie	Year	Rating	Time (min)	Genre	Box office ($)
Avatar	2009	PG-13	162	Action	2,783,918,982
Titanic	1997	PG-13	194	Drama	2,207,615,668
Star Wars: The Force Awakens	2015	PG-13	136	Adventure	2,040,375,795
Jurassic World	2015	PG-13	124	Action	1,669,164,161
Marvel's The Avengers	2012	PG-13	142	Action	1,519,479,547
Furious 7	2015	PG-13	137	Action	1,516,246,709
The Avengers: Age of Ultron	2015	PG-13	141	Action	1,404,705,868
Harry Potter and the Deathly Hallows: Part 2	2011	PG-13	130	Fantasy	1,328,111,219
Frozen	2013	PG	108	Animation	1,254,512,386
Iron Man 3	2013	PG-13	129	Action	1,172,805,920

4. **Skyscrapers** Here is some information about the tallest buildings in the world as of February 2017. Identify the individuals and variables in this data set. Classify each variable as categorical or quantitative.

Building	Country	Height (m)	Floors	Use	Year completed
Burj Khalifa	United Arab Emirates	828.0	163	Mixed	2010
Shanghai Tower	China	632.0	121	Mixed	2014
Makkah Royal Clock Tower Hotel	Saudi Arabia	601.0	120	Hotel	2012
Ping An Finance Center	China	599.0	115	Mixed	2016
Lotte World Tower	South Korea	554.5	123	Mixed	2016
One World Trade Center	United States	541.0	104	Office	2013
Taipei 101	Taiwan	509.0	101	Office	2004
Shanghai World Financial Center	China	492.0	101	Mixed	2008
International Commerce Center	China	484.0	118	Mixed	2010
Petronas Tower 1	Malaysia	452.0	88	Office	1998

5. **Protecting wood** What measures can be taken, especially when restoring historic wooden buildings, to help wood surfaces resist weathering? In a study of this question, researchers prepared wooden panels and then exposed them to the weather. Some of the variables recorded were type of wood (yellow poplar, pine, cedar); type of water repellent (solvent-based, water-based); paint thickness (millimeters); paint color (white, gray, light blue); weathering time (months). Classify each variable as categorical or quantitative.

6. **Medical study variables** Data from a medical study contain values of many variables for each subject in the study. Some of the variables recorded were gender (female or male); age (years); race (Asian, Black, White, or other); smoker (yes or no); systolic blood pressure (millimeters of mercury); level of calcium in the blood (micrograms per milliliter). Classify each variable as categorical or quantitative.

7. **College life** A college admissions office collects data from each incoming freshman on several quantitative variables: distance from their home to campus, number of siblings, how many books they have read in the past month, and how long it took them to complete an online survey. Classify each variable as discrete or continuous.

8. **Social media** A social media company records data from each of its users on several quantitative variables: time spent on the site, how many times they visited the site, number of likes received, and how long since they created a member profile. Classify each variable as discrete or continuous.

Multiple Choice: *Select the best answer.*

Exercises 9 and 10 refer to the following setting. At the Census Bureau website www.census.gov, you can view detailed data collected by the American Community Survey. The following table includes data for 10 people chosen at random from the more than 1 million people in households contacted by the survey. "School" gives the highest level of education completed.

Weight (lb)	Age (years)	Travel to work (min)	School	Gender	Income last year ($)
187	66	0	Ninth grade	1	24,000
158	66	n/a	High school grad	2	0
176	54	10	Assoc. degree	2	11,900
339	37	10	Assoc. degree	1	6000
91	27	10	Some college	2	30,000
155	18	n/a	High school grad	2	0
213	38	15	Master's degree	2	125,000
194	40	0	High school grad	1	800
221	18	20	High school grad	1	2500
193	11	n/a	Fifth grade	1	0

9. The individuals in this data set are

(a) households. (b) people. (c) adults.

(d) 120 variables. (e) columns.

10. This data set contains

(a) 7 variables, 2 of which are categorical.

(b) 7 variables, 1 of which is categorical.

(c) 6 variables, 2 of which are categorical.

(d) 6 variables, 1 of which is categorical.

(e) None of these.