



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)*, Fifth Edition, is full of data. Each set of data has some brief background to help you understand what the data say. We deliberately chose contexts and data sets in the examples and exercises to pique your interest.

TPS 5e 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 5e* has to offer. You can learn a lot by reading the text, but you will develop deeper understanding by doing Activities and Data Explorations and answering the Check Your Understanding questions along the way. The walkthrough guide on pages xiv–xx 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[®] exam. Recycle and Review exercises at the end of each exercise set involve material you studied in previous sections.
- **Chapter Review Exercises** consist of free-response questions aligned to specific learning objectives from the chapter. Go through the list of learning objectives summarized in the Chapter Review and be sure you can say “I can do that” to each item. 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 10 to 12 multiple-choice questions and three free-response problems, very much in the style of the AP[®] exam.
- Finally, the **Cumulative AP[®] Practice Tests** after Chapters 4, 7, 10, and 12 provide challenging, cumulative multiple-choice and free-response questions like ones 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 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. Like the previous four editions, *TPS 5e* is organized to closely follow the AP[®] Statistics Course Description. Every item on the College Board’s “Topic Outline” is covered thoroughly in the text. Look inside the front cover for a detailed alignment guide. The few topics in the book that go beyond the AP[®] syllabus are marked with an asterisk (*).

Most importantly, *TPS 5e* is designed to prepare you for the AP[®] Statistics exam. The entire author team has been involved in the AP[®] Statistics program since its early days. We have more than 80 years’ combined experience teaching introductory statistics and more than 30 years’ combined experience grading the AP[®] exam! Two of us (Starnes and Tabor) have served as Question Leaders for several years, helping to write scoring rubrics for free-response questions. Including our Content Advisory Board and Supplements Team (page vii), we have two former Test Development Committee members and 11 AP[®] exam Readers.

TPS 5e will help you get ready for the AP[®] Statistics exam throughout the course by:

- **Using terms, notation, formulas, and tables consistent with those found on the AP[®] 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[®] 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 5e*. For example, the four-step State-Plan-Do-Conclude process that we use to complete inference problems in Chapters 8 through 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 and in some Technology Corners as “on-the-spot” reminders of common mistakes and how to avoid them. These tips are collected and summarized in Appendix A.
- **Providing hundreds of AP[®]-style exercises throughout the book.** We even added a new kind of problem just prior to each Chapter Review, called 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 Web site (www.whfreeman.com/tps5e). Then you can score your own response using a rubric provided by your teacher.

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[®] exam.

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

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

Scan the margins for the purple notes, which represent the “voice of the teacher” giving helpful hints for being successful in the course.

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

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

3.1 Scatterplots and Correlation

WHAT YOU WILL LEARN By the end of the section, you should be able to:

- Identify explanatory and response variables in situations where one variable helps to explain or influences the other.
- 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 outliers in a scatterplot.
- Interpret the correlation.
- Understand the basic properties of correlation, including how the correlation is influenced by outliers.
- Use technology to calculate correlation.
- Explain why association does not imply causation.

Often, using the regression line to make a prediction for $x = 0$ is an extrapolation. That's why the y intercept isn't always statistically meaningful.

DEFINITION: Extrapolation

Extrapolation is the use of a regression line for prediction far outside the interval of values of the explanatory variable x used to obtain the line. Such predictions are often not accurate.

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.*



Take note of the green DEFINITION boxes that explain important vocabulary. Flip back to them to review key terms and their definitions.

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

HOW TO MAKE A SCATTERPLOT

- Decide which variable should go on each axis.
- Label and scale your axes.
- Plot individual data values.

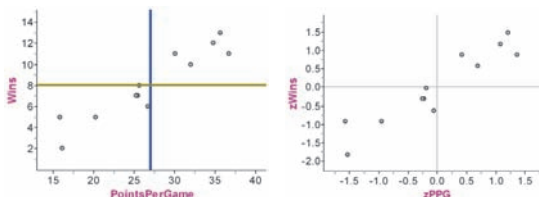
Make connections and deepen your understanding by reflecting on the questions asked in THINK ABOUT IT passages.

AP[®] EXAM TIP The formula sheet for the AP[®] exam uses different notation for these equations: $b_1 = r \frac{s_y}{s_x}$ and $b_0 = \bar{y} - b_1 \bar{x}$. That's because the least-squares line is written as $\hat{y} = b_0 + b_1 x$. We prefer our simpler versions without the subscripts!

THINK ABOUT IT

What does correlation measure? The Fathom screen shots below provide more detail. At the left is a scatterplot of the SEC football data with two lines added—a vertical line at the group's mean points per game and a horizontal line at the mean number of wins of the group. Most of the points fall in the upper-right “quadrants” of the graph. That is, teams with above-average points per game tend to have above-average numbers of wins, and teams with below-average points per game tend to have numbers of wins that are below average. This confirms the positive association between the variables.

Below on the right is a scatterplot of the standardized scores. To get this graph, we transformed both the x - and the y -values by subtracting their mean and dividing by their standard deviation. As we saw in Chapter 2, standardizing a data set converts the mean to 0 and the standard deviation to 1. That's why the vertical and horizontal lines in the right-hand graph are both at 0.



Notice that all the products of the standardized values will be positive—not surprising, considering the strong positive association between the variables. What if there was a negative association between two variables? Most of the points would be in the upper-left and lower-right “quadrants” and their z -score products would be negative, resulting in a negative correlation.

LEARN STATISTICS BY *DOING* STATISTICS

ACTIVITY | Reaching for Chips

MATERIALS:

200 colored chips, including 100 of the same color; large bag or other container



Before class, your teacher will prepare a population of having the same color (say, red). The parameter is the chips in the population: $p = 0.50$. In this Activity, variability by taking repeated random samples of size

1. After your teacher has mixed the chips thoroughly should take a sample of 20 chips and note the sample. When finished, the student should return all the chips and pass the bag to the next student.

Note: If your class has fewer than 25 students, have samples.

2. Each student should record the \hat{p} -value in a chart value on a class dotplot. Label the graph scale from spaced 0.05 units apart.

3. Describe what you see: shape, center, spread, and usual features.

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.

DATA EXPLORATIONS ask you to play the role of data detective. Your goal is to answer a puzzling, real-world question by examining data graphically and numerically.

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

ACTIVITY | I'm a Great Free-Throw Shooter!

MATERIALS:

Computer with Internet access and projection capability



A basketball player claims to make 80% of the free throws that he attempts. We think he might be exaggerating. To test this claim, we'll ask him to shoot some free throws—virtually—using *The Reasoning of a Statistical Test* applet at the book's Web site.

1. Go to www.whfreeman.com/tps5e and launch the applet.



2. Set the applet to take 25 shots. Click "Shoot." How many of the 25 shots did the player make? Do you have enough data to decide whether the player's claim is valid?

3. Click "Shoot" again for 25 more shots. Keep doing this until you are convinced *either* that the player makes less than 80% of his shots *or* that the player's claim is true. How large a sample of shots did you need to make your decision?

4. Click "Show true probability" to reveal the truth. Was your conclusion correct?

5. If time permits, choose a new shooter and repeat Steps 2 through 4. Is it easier to tell that the player is exaggerating when his actual proportion of free throws made is closer to 0.8 or farther from 0.8?

DATA EXPLORATION The SAT essay: Is longer better?

Following the debut of the new SAT Writing test in March 2005, Dr. Les Perelman from the Massachusetts Institute of Technology stirred controversy by reporting, "It appeared to me that regardless of what a student wrote, the longer the essay, the higher the score." He went on to say, "I have never found a quantifiable predictor in 25 years of grading that was anywhere as strong as this one. If you just graded them based on length without ever reading them, you'd be right over 90 percent of the time."³ The table below shows the data that Dr. Perelman used to draw his conclusions.⁴

	Length of essay and score for a sample of SAT essays										
Words:	460	422	402	365	357	278	236	201	168	156	133
Score:	6	6	5	5	6	5	4	4	4	3	2
Words:	114	108	100	403	401	388	320	258	236	189	128
Score:	2	1	1	5	6	6	5	4	4	3	2
Words:	67	697	387	355	337	325	272	150	135		
Score:	1	6	6	5	5	4	4	2	3		

Does this mean that if students write a lot, they are guaranteed high scores? Carry out your own analysis of the data. How would you respond to each of Dr. Perelman's claims?



CHECK YOUR UNDERSTANDING

Identify the explanatory and response variables in each setting.

- How does drinking beer affect the level of alcohol in people's blood? The legal limit for driving in all states is 0.08%. In a study, adult volunteers drank different numbers of cans of beer. Thirty minutes later, a police officer measured their blood alcohol levels.
- The National Student Loan Survey provides data on the amount of debt for recent college graduates, their current income, and how stressed they feel about college debt. A sociologist looks at the data with the goal of using amount of debt and income to explain the stress caused by college debt.

EXAMPLES: Model statistical problems and how to solve them

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 actually specify 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 but just observe both variables, there may or may not be explanatory and response variables. Whether there are depends on how you plan to use the data.



EXAMPLE

Linking SAT Math and Critical Reading Scores

Explanatory or response?


Julie asks, “Can I predict a state’s mean SAT Math score if I know its mean SAT Critical Reading score?” Jim wants to know how the mean SAT Math and Critical Reading scores this year in the 50 states are related to each other.

PROBLEM: For each student, identify the explanatory variable and the response variable if possible.

SOLUTION: Julie is treating the mean SAT Critical Reading score as the explanatory variable and the mean SAT Math score as the response variable. Jim is simply interested in exploring the relationship between the two variables. For him, there is no clear explanatory or response variable.

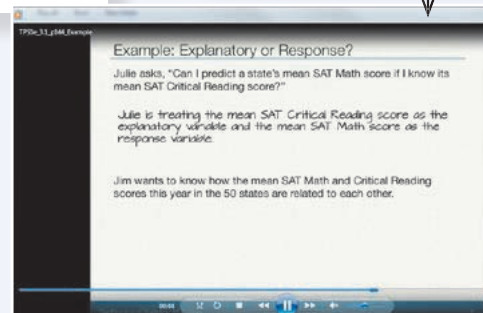
For Practice Try Exercise 1

Read through each EXAMPLE, and then try out the concept yourself by working the FOR PRACTICE 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[®] teachers. The video guides you through each step in the example and solution and gives you extra help when you need it.

The red number box next to the exercise directs you back to the page in the section where the model example appears.

1. Coral reefs How sensitive to changes in water temperature are coral reefs? To find out, measure the growth of corals in aquariums where the water temperature is controlled at different levels. Growth is measured by weighing the coral before and after the experiment. What are the explanatory and response variables? Are they categorical or quantitative?



EXAMPLE

Gesell Scores

STEP 4

Putting it all together

Does the age at which a child begins to talk predict a later score on a test of mental ability? A study of the development of young children recorded the age in months at which each of 21 children spoke their first word and their Gesell Adaptive Score, the result of an aptitude test taken much later.¹⁶ The data appear in the table below, along with a scatterplot, residual plot, and computer output. Should we use a linear model to predict a child’s Gesell score from his or her age at first word? If so, how accurate will our predictions be?



Age (months) at first word and Gesell score								
CHILD	AGE	SCORE	CHILD	AGE	SCORE	CHILD	AGE	SCORE
1	15	95	8	11	100	15	11	102
2	26	71	9	8	104	16	10	100
3	10	83	10	20	94	17	12	105
4	9	91	11	7	113	18	42	57
5	15	102	12	9	96	19	17	121
6	20	87	13	10	83	20	11	86
7	18	93	14	11	84	21	10	100

4-STEP EXAMPLES: By reading the 4-Step Examples and mastering the special “State-Plan-Do-Conclude” framework, you can develop good problem-solving skills and your ability to tackle more complex problems like those on the AP[®] exam.

EXERCISES: Practice makes perfect!

Start by reading the SECTION SUMMARY to be sure that you understand the key concepts.

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 problems test the same skill or concept. If you answer an assigned problem incorrectly, try to figure out your mistake. Then see if you can solve the paired exercise.

Look for icons that appear next to selected problems. They will guide you to

- an Example that models the problem.
- videos that provide step-by-step instructions for solving the problem.
- earlier sections on which the problem draws (here, Section 2.2).
- examples with the 4-Step State-Plan-Do-Conclude way of solving problems.

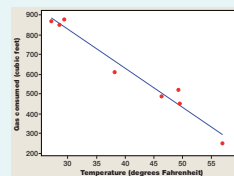
Section 3.2 Summary

- A **regression line** is a straight line that describes how a response variable y changes as an explanatory variable x changes. You can use a regression line to **predict** the value of y for any value of x by substituting this x into the equation of the line.
- The **slope** b of a regression line $\hat{y} = a + bx$ is the rate at which the predicted response \hat{y} changes along the line as the explanatory variable x changes. Specifically, b is the *predicted* change in y when x increases by 1 unit.
- The **y intercept** a of a regression line $\hat{y} = a + bx$ is the predicted response \hat{y} when the explanatory variable x equals 0. This prediction is of no statistical use unless x can actually take values near 0.

Section 3.2 Exercises

35. **What's my line?** You use the same bar of soap to shower each morning. The bar weighs 80 grams when it is new. Its weight goes down by 6 grams per day on average. What is the equation of the regression line for predicting weight from days of use?
36. **What's my line?** An eccentric professor believes that a child with IQ 100 should have a reading test score of 50 and predicts that reading score should increase by 1 point for every additional point of IQ. What is the equation of the professor's regression line for predicting reading score from IQ?
37. **Gas mileage** We expect a car's highway gas mileage to be related to its city gas mileage. Data for all 1198 vehicles in the government's recent *Fuel Economy Guide* give the regression line: predicted highway mpg = $4.62 + 1.109(\text{city mpg})$.
- What's the slope of this line? Interpret this value in context.
 - What's the y intercept? Explain why the value of the intercept is not statistically meaningful.
 - Find the predicted highway mileage for a car that gets 16 miles per gallon in the city.
38. **IQ and reading scores** Data on the IQ test scores and reading test scores for a group of fifth-grade children give the following regression line: predicted reading score = $-33.4 + 0.882(\text{IQ score})$.
- What's the slope of this line? Interpret this context.
 - What's the y intercept? Explain why the intercept is not statistically meaningful.
 - Find the predicted reading score for a child with an IQ score of 90.

in Joan's midwestern home. The figure below shows the original scatterplot with the least-squares line added. The equation of the least-squares line is $\hat{y} = 1425 - 19.87x$.



- Identify the slope of the line and explain what it means in this setting.
 - Identify the y intercept of the line. Explain why it's risky to use this value as a prediction.
 - Use the regression line to predict the amount of natural gas Joan will use in a month with an average temperature of 30°F.
41. **Acid rain** Refer to Exercise 39. Would it be appropriate to use the regression line to predict pH after 1000 months? Justify your answer.

39. **Acid rain** Researchers studying acid rain measured the acidity of precipitation in a Colorado area for 150 consecutive weeks. Acidity is measured by pH. Lower pH values show higher acidity. Researchers observed a linear pattern over time. They reported that the regression line $\hat{pH} = 5.43 - 0.0053(\text{weeks})$ fit the data well.¹⁹

79. **In my Chevrolet (2.2)** The Chevrolet Malibu with a four-cylinder engine has a combined gas mileage of 25 mpg. What percent of all vehicles have worse gas mileage than the Malibu?

67. **Beavers and beetles** Do beavers benefit beetles? Researchers laid out 23 circular plots, each 4 meters in diameter, in an area where beavers were cutting down cottonwood trees. In each plot, they counted the number of stumps from trees cut by beavers and the number of clusters of beetle larvae. Ecologists think that the new sprouts from stumps are more tender than other cottonwood growth, so that beetles prefer them.

REVIEW and PRACTICE for quizzes and tests

Chapter 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 associations between two quantitative variables. To describe a scatterplot, be sure to discuss four characteristics: direction, form, strength, and outliers. The direction of an association might be positive, negative, or neither. The form of an association can be linear or nonlinear. An association is strong if it closely follows a specific form. Finally, outliers are any points that clearly fall outside the pattern of the rest of the data.

The correlation r is a numerical summary that describes the direction and strength of a linear association. When $r > 0$, the association is positive, and when $r < 0$, the association is negative. The correlation will always take values between -1 and 1 , with $r = -1$ and $r = 1$ indicating a perfectly linear relationship. Strong linear associations have correlations near 1 or -1 , while weak linear relationships have correlations near 0 . However, it is

possible to determine the form of an association from only the correlation. Strong nonlinear relationships can have a correlation close to 1 or a correlation close to 0 , depending on the association. You also learned that outliers can greatly affect the value of the correlation and that correlation does not imply causation. That is, we can't assume that changes in one variable cause changes in the other variable, just because they have a correlation close to 1 or -1 .

Section 3.2: Least-Squares Regression

In this section, you learned how to use least-squares regression lines as models for relationships between variables that have a linear association. It is important to understand the difference between the actual data and the model used to describe the data. For example, when you are interpreting the slope of a least-squares regression

Review the CHAPTER SUMMARY to be sure that you understand the key concepts in each section.

Use the WHAT DID YOU LEARN? table to guide you to model examples and exercises to verify your mastery of each LEARNING OBJECTIVE.

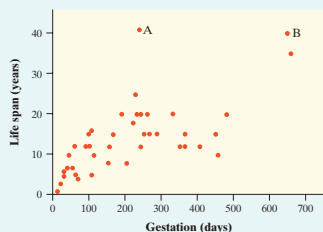
What Did You Learn?

Learning Objective	Section	Related Example on Page(s)	Relevant Chapter Review Exercise(s)
Identify explanatory and response variables in situations where one variable helps to explain or influences the other.	3.1	144	R3.4
Make a scatterplot to display the relationship between two quantitative variables.	3.1	145, 148	R3.4
Describe the direction, form, and strength of a relationship displayed in a scatterplot and recognize outliers in a scatterplot.	3.1	147, 148	R3.1
Interpret the correlation.	3.1	152	R3.3, R3.4
Understand the basic properties of correlation, including how the correlation is influenced by outliers.	3.1	152, 156, 157	R3.1, R3.2
Use technology to calculate correlation.	3.1	Activity on 152, 171	R3.4
Explain why association does not imply causation.	3.1	Discussion on 156, 190	R3.6
Interpret the slope and y -intercept of a least-squares regression line.	3.2	166	R3.2, R3.4
Use the least-squares regression line to predict y for a given x .	3.2	167, Discussion on 168	R3.2, R3.4, R3.5
Explain the dangers of extrapolation.	3.2	(for extrapolation)	R3.2, R3.4, R3.5
Calculate and interpret residuals.	3.2	169	R3.3, R3.4
Explain the concept of least squares.	3.2	Discussion on 169	R3.5
Determine the equation of a least-squares regression line using technology or computer output.	3.2	Technology Corner on 171, 181	R3.3, R3.4
Construct and interpret residual plots to assess whether a linear model is appropriate.	3.2	Discussion on 175, 180	R3.3, R3.4
Interpret the standard deviation of the residuals and r^2 and use these values to assess how well the least-squares regression line models the relationship between two variables.	3.2	180	R3.3, R3.5
		Discussion on 188	R3.1
		183	R3.5

Chapter 3 Chapter 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.³⁰

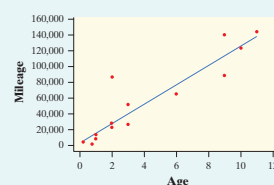


(a) Describe the association shown in the scatterplot.

R3.3 Stats teachers' cars A random sample of AP[®] Statistics teachers was asked to report the age (in years) and mileage of their primary vehicles. A scatterplot of the data, a least-squares regression printout, and a residual plot are provided below.

Predictor	Coef	SE Coef	T	P
Constant	3704	8268	0.45	0.662
Age	12188	1492	8.17	0.000

S = 20870.5 R-Sq = 83.7% R-Sq(adj) = 82.4%



Tackle the CHAPTER REVIEW EXERCISES for practice in solving problems that test concepts from throughout the chapter.

and the AP[®] Exam

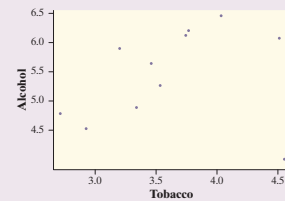
Each chapter concludes with an AP[®] STATISTICS PRACTICE TEST. This test includes about 10 AP[®]-style multiple-choice questions and 3 free-response questions.

Chapter 3 AP[®] Statistics Practice Test

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

- T3.1** A school guidance counselor examines the number of extracurricular activities that students do 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 zero." A correct interpretation of this statement would be that
- active students tend to be students with poor grades, and vice versa.
 - students with good grades tend to be students who are not involved in many extracurricular activities, and vice versa.
 - students involved in many extracurricular activities are just as likely to get good grades as bad grades; the same is true for students involved in few extracurricular activities.
 - there is no linear relationship between number of activ-

alcoholic beverages for each of 11 regions in Great Britain was recorded. A scatterplot of spending on alcohol versus spending on tobacco is shown below. Which of the following statements is true?



- The observation (4.5, 6.0) is an outlier.
- There is clear evidence of a negative association between alcohol and tobacco.
- The equation of the least-squares line for this plot is approximately $\hat{y} = 10 - 2x$.
- The correlation for these data is $r = 0.99$.
- The observation in the lower-right corner of the plot is influential for the least-squares line.

Cumulative AP[®] Practice Test 1

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

- AP1.1** You look at real estate ads for houses in Sarasota, Florida. Many houses range from \$200,000 to \$400,000 in price. 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?
- The distribution is most likely skewed to the left, and the mean is greater than the median.

- AP1.4** For a certain experiment, the available experimental units are eight rats, of which four are female (F1, F2, F3, F4) and four are male (M1, M2, M3, M4). There are to be four treatment groups, A, B, C, and D. If a randomized block design is used, with the experimental units blocked by gender, which of the following assignments of treatments is impossible?

Section II: Free Response Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

- AP1.15** The manufacturer of exercise machines for fitness centers has designed two new elliptical machines that are meant to increase cardiovascular fitness. The two machines are being tested on 30 volunteers at a fitness center near the company's headquarters. The volunteers are randomly assigned to one of the machines and use it daily for two months. A measure of cardiovascular fitness is administered at the start of the experiment and

the two machines. Note that higher scores indicate larger gains in fitness.

Machine A	Machine B
0	2
54	1
876320	2
97411	3
4	2
2	1

Four CUMULATIVE AP[®] TESTS simulate the real exam. They are placed after Chapters 4, 7, 10, and 12. The tests expand in length and content coverage from the first through the fourth.

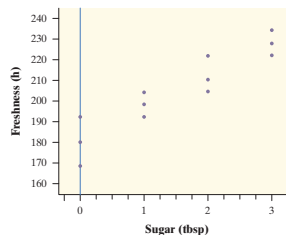
Learn how to answer free-response questions successfully by working the FRAPPY! THE FREE RESPONSE AP[®] PROBLEM, YAY! that comes just before the Chapter Review in every chapter.

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.

and observed how many hours each flower continued to look fresh. A scatterplot of the data is shown below.



- Briefly describe the association shown in the scatterplot.
- 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.

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

- Calculate and interpret the residual for the flower that had 2 tablespoons of sugar and looked fresh for 204 hours.
- 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 Web site (www.whfreeman.com/tps5e). 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 to improve your own score.

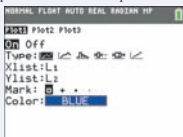
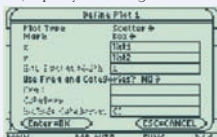
Use TECHNOLOGY to discover and analyze

Use technology as a tool for discovery and analysis. TECHNOLOGY CORNERS give step-by-step instructions for using the TI-83/84 and TI-89 calculator. Instructions for the TI-Nspire are in an end-of-book appendix. HP Prime instructions are on the book's Web site and in the e-Book.

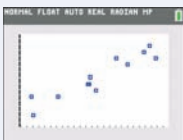
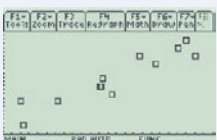
7. TECHNOLOGY CORNER **SCATTERPLOTS ON THE CALCULATOR**
 TI-Nspire instructions in Appendix B; HP Prime instructions on the book's Web site

Making scatterplots with technology is much easier than constructing them by hand. We'll use the SEC football data from page 146 to show how to construct a scatterplot on a TI-83/84 or TI-89.

- Enter the data values into your lists. Put the points per game in L1/list1 and the number of wins in L2/list2.
- Define a scatterplot in the statistics plot menu (press **2ND** on the TI-89). Specify the settings shown below.

- Use ZoomStat (ZoomData on the TI-89) to obtain a graph. The calculator will set the window dimensions automatically by looking at the values in L1/list1 and L2/list2.

Notice that there are no scales on the axes and that the axes are not labeled. If you copy a scatterplot from your calculator onto your paper, make sure that you scale and label the axes.

AP® EXAM TIP If you are asked to make a scatterplot on a free-response question, be sure to label and scale both axes. *Don't* just copy an unlabeled calculator graph directly onto your paper.

You can access video instructions for the Technology Corners through the e-Book or on the book's Web site.



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

3.2 TECHNOLOGY CORNERS
 TI-Nspire Instructions in Appendix B; HP Prime instructions on the book's Web site

8. Least-squares regression lines on the calculator	page 171
9. Residual plots on the calculator	page 175

59. Merlins breeding Exercise 13 (page 160) gives data on the number of breeding pairs of merlins in an isolated area in each of seven years and the percent of males who returned the next year. The data show that the percent returning is lower after successful breeding seasons and that the relationship is roughly linear. The figure below shows Minitab regression output for these data.



- What is the equation of the least-squares regression line for predicting the percent of males that return from the number of breeding pairs? Use the equation to predict the percent of returning males after a season with 30 breeding pairs.
- What percent of the year-to-year variation in percent of returning males is accounted for by the straight-line relationship with number of breeding pairs the previous year?

Other types of software displays, including Minitab, Fathom, and applet screen captures, appear throughout the book to help you learn to read and interpret many different kinds of output.



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? Should a poker player go “all in” with pocket aces? 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

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 decisions 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.

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?

You shouldn’t. The people who took the trouble to write 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 three 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.

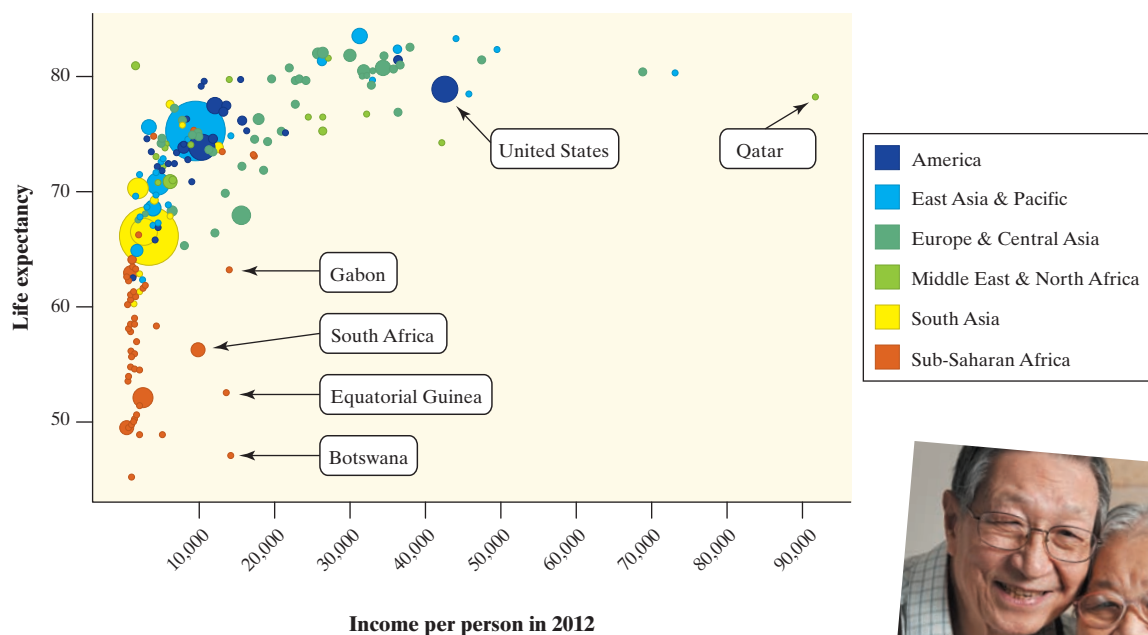
Always Plot Your Data

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 is often more instructive than a bunch of numbers.*

Do people live longer in wealthier countries?

The Gapminder Web site, www.gapminder.org, provides loads of data on the health and well-being of the world’s inhabitants. The graph on the next pages 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 in years for one country.

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 2012.



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?

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 they asked 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 a language for talking about uncertainty that is understood by statistically literate people everywhere.

