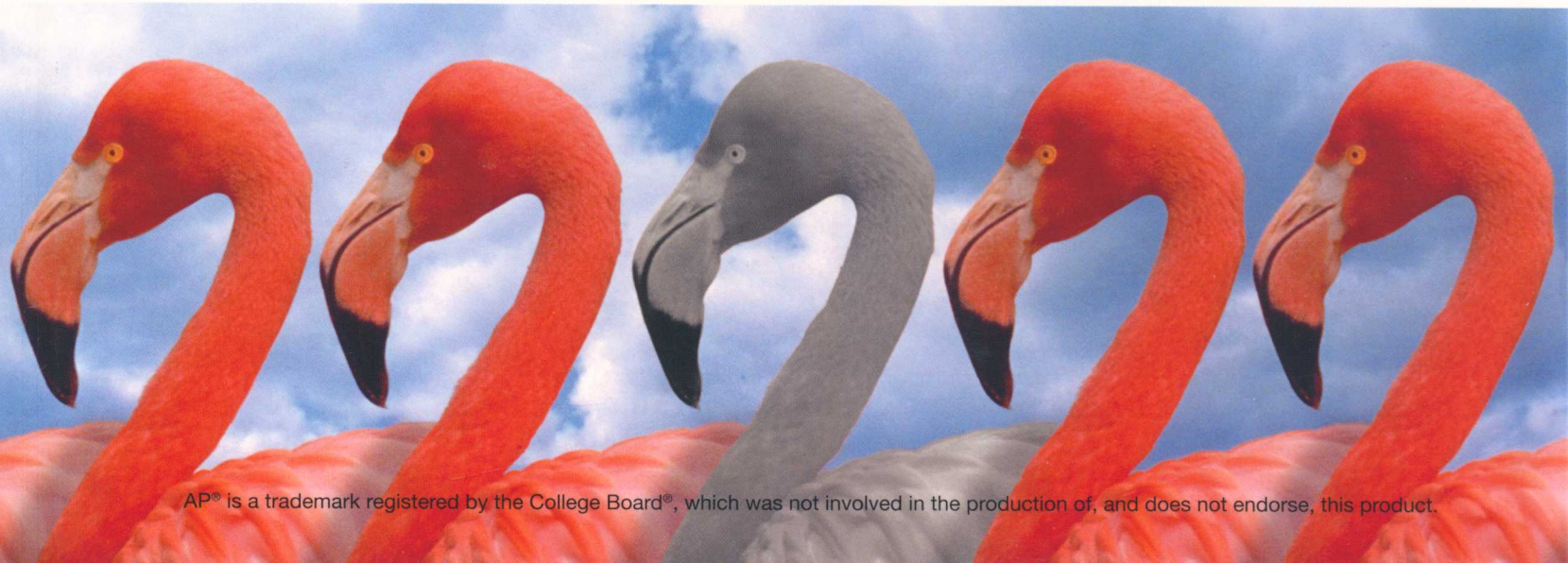
FOR THE AP® EXAM

The PRACTICE of STATISTICS

FIFTH EDITION

STARNES • TABOR • YATES • MOORE



For the **AP®** Exam

The Practice of Statistics

FIFTH EDITION

Publisher: Ann Heath
Assistant Editor: Enrico Bruno
Editorial Assistant: Matt Belford
Development Editor: Donald Gecewicz
Executive Marketing Manager: Cindi Weiss
Photo Editor: Cecilia Varas
Photo Researcher: Julie Tesser
Art Director: Diana Blume
Cover Designers: Diana Blume, Rae Grant
Text Designer: Patrice Sheridan
Cover Image: Joseph Devenney/Getty Images
Senior Project Editor: Vivien Weiss
Illustrations: Network Graphics
Production Manager: Susan Wein
Composition: Preparé
Printing and Binding: RR Donnelley

Library of Congress Control Number: 2013949503
ISBN-13: 978-1-4641-0873-0
ISBN-10: 1-4641-0873-0

First printing 2014

Printed in the United States of America

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

Using *The Practice of Statistics*, Fifth Edition, for Advanced Placement (AP®) Statistics

(The percents in parentheses reflect coverage on the AP® exam.)

Topic Outline for AP® Statistics from the College Board's *AP® Statistics Course Description*	*The Practice of Statistics*, 5th ed. Chapter and Section references
I. Exploring data: describing patterns and departures from patterns (20%–30%)	
A. Constructing and interpreting graphical displays of distributions of univariate data (dotplot, stemplot, histogram, cumulative frequency plot)	**Dotplot, stemplot, histogram 1.2; Cumulative frequency plot 2.1**
1. Center and spread	1.2
2. Clusters and gaps	1.2
3. Outliers and unusual features	1.2
4. Shape	1.2
B. Summarizing distributions of univariate data	**1.3 and 2.1**
1. Measuring center: median, mean	1.3
2. Measuring spread: range, interquartile range, standard deviation	1.3
3. Measuring position: quartiles, percentiles, standardized scores (*z*-scores)	Quartiles 1.3; percentiles and *z*-scores 2.1
4. Using boxplots	1.3
5. The effect of changing units on summary measures	2.1
C. Comparing distributions of univariate data (dotplots, back-to-back stemplots, parallel boxplots)	**Dotplots and stemplots 1.2; boxplots 1.3**
1. Comparing center and spread	1.2 and 1.3
2. Comparing clusters and gaps	1.2 and 1.3
3. Comparing outliers and unusual features	1.2 and 1.3
4. Comparing shape	1.2 and 1.3
D. Exploring bivariate data	**Chapter 3 and Section 12.2**
1. Analyzing patterns in scatterplots	3.1
2. Correlation and linearity	3.1
3. Least-squares regression line	3.2
4. Residual plots, outliers, and influential points	3.2
5. Transformations to achieve linearity: logarithmic and power transformations	12.2
E. Exploring categorical data	**Sections 1.1, 5.2, 5.3**
1. Frequency tables and bar charts	1.1 (we call them bar graphs)
2. Marginal and joint frequencies for two-way tables	Marginal 1.1; joint 5.2
3. Conditional relative frequencies and association	1.1 and 5.3
4. Comparing distributions using bar charts	1.1
II. Sampling and experimentation: planning and conducting a study (10%–15%)	
A. Overview of methods of data collection	**Sections 4.1 and 4.2**
1. Census	4.1
2. Sample survey	4.1
3. Experiment	4.2
4. Observational study	4.2
B. Planning and conducting surveys	**Section 4.1**
1. Characteristics of a well-designed and well-conducted survey	4.1
2. Populations, samples, and random selection	4.1
3. Sources of bias in sampling and surveys	4.1
4. Sampling methods, including simple random sampling, stratified random sampling, and cluster sampling	4.1
C. Planning and conducting experiments	**Section 4.2**
1. Characteristics of a well-designed and well-conducted experiment	4.2
2. Treatments, control groups, experimental units, random assignments, and replication	4.2
3. Sources of bias and confounding, including placebo effect and blinding	4.2
4. Completely randomized design	4.2
5. Randomized block design, including matched pairs design	4.2
D. Generalizability of results and types of conclusions that can be drawn from observational studies, experiments, and surveys	**Section 4.3**

Using *The Practice of Statistics*, Fifth Edition, for Advanced Placement (AP®) Statistics

(The percents in parentheses reflect coverage on the AP® exam.)

Topic Outline for AP® Statistics from the College Board's *AP® Statistics Course Description*	*The Practice of Statistics*, 5th ed. Chapter and Section references
III. Anticipating patterns: exploring random phenomena using probability and simulation (20%–30%)	
A. Probability	**Chapters 5 and 6**
1. Interpreting probability, including long-run relative frequency interpretation	5.1
2. "Law of large numbers" concept	5.1
3. Addition rule, multiplication rule, conditional probability, and independence	Addition rule 5.2; other three topics 5.3
4. Discrete random variables and their probability distributions, including binomial and geometric	Discrete 6.1; Binomial and geometric 6.3
5. Simulation of random behavior and probability distributions	5.1
6. Mean (expected value) and standard deviation of a random variable, and linear transformation of a random variable	Mean and standard deviation 6.1; Linear transformation 6.2
B. Combining independent random variables	**Section 6.2**
1. Notion of independence versus dependence	6.2
2. Mean and standard deviation for sums and differences of independent random variables	6.2
C. The Normal distribution	**Section 2.2**
1. Properties of the Normal distribution	2.2
2. Using tables of the Normal distribution	2.2
3. The Normal distribution as a model for measurements	2.2
D. Sampling distributions	**Chapter 7; Sections 8.3, 10.1, 10.2, 11.1**
1. Sampling distribution of a sample proportion	7.2
2. Sampling distribution of a sample mean	7.3
3. Central limit theorem	7.3
4. Sampling distribution of a difference between two independent sample proportions	10.1
5. Sampling distribution of a difference between two independent sample means	10.2
6. Simulation of sampling distributions	7.1
7. *t* distribution	8.3
8. Chi-square distribution	11.1
IV. Statistical inference: estimating population parameters and testing hypotheses (30%–40%)	
A. Estimation (point estimators and confidence intervals)	**Chapter 8 plus parts of Sections 9.3, 10.1, 10.2, 12.1**
1. Estimating population parameters and margins of error	8.1
2. Properties of point estimators, including unbiasedness and variability	8.1
3. Logic of confidence intervals, meaning of confidence level and confidence intervals, and properties of confidence intervals	8.1
4. Large-sample confidence interval for a proportion	8.2
5. Large-sample confidence interval for a difference between two proportions	10.1
6. Confidence interval for a mean	8.3
7. Confidence interval for a difference between two means (unpaired and paired)	Paired 9.3; unpaired 10.2
8. Confidence interval for the slope of a least-squares regression line	12.1
B. Tests of significance	**Chapters 9 and 11 plus parts of Sections 10.1, 10.2, 12.1**
1. Logic of significance testing, null and alternative hypotheses; *P*-values; one-and two-sided tests; concepts of Type I and Type II errors; concept of power	9.1; power in 9.2
2. Large-sample test for a proportion	9.2
3. Large-sample test for a difference between two proportions	10.1
4. Test for a mean	9.3
5. Test for a difference between two means (unpaired and paired)	Paired 9.3; unpaired 10.2
6. Chi-square test for goodness of fit, homogeneity of proportions, and independence (one- and two-way tables)	Chapter 11
7. Test for the slope of a least-squares regression line	12.1

For the **AP®** Exam

The Practice of Statistics

FIFTH EDITION

Daren S. Starnes
The Lawrenceville School

Josh Tabor
Canyon del Oro High School

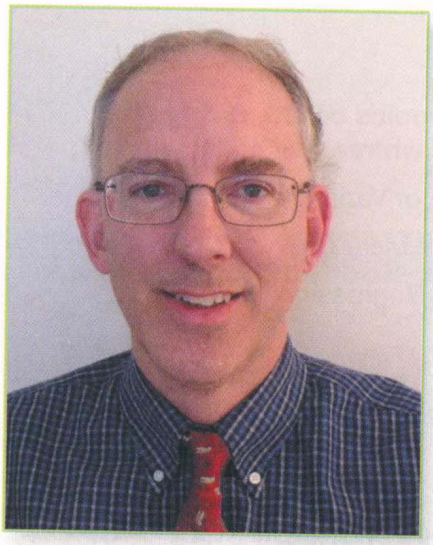
Daniel S. Yates
Statistics Consultant

David S. Moore
Purdue University

bfw Freeman

W. H. Freeman and Company/BFW
New York

About the Authors



DAREN S. STARNES is Mathematics Department Chair and holds the Robert S. and Christina Seix Dow Distinguished Master Teacher Chair in Mathematics at The Lawrenceville School near Princeton, New Jersey. He earned his MA in Mathematics from the University of Michigan and his BS in Mathematics from the University of North Carolina at Charlotte. Daren is also an alumnus of the North Carolina School of Science and Mathematics. Daren has led numerous one-day and weeklong AP® Statistics institutes for new and experienced AP® teachers, and he has been a Reader, Table Leader, and Question Leader for the AP® Statistics exam since 1998. Daren is a frequent speaker at local, state, regional, national, and international conferences. For two years, he served as coeditor of the Technology Tips column in the NCTM journal *The Mathematics Teacher*. From 2004 to 2009, Daren served on the ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability (which he chaired in 2009). While on the committee, he edited the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) pre-K–12 report and coauthored (with Roxy Peck) *Making Sense of Statistical Studies*, a capstone module in statistical thinking for high school students. Daren is also coauthor of the popular text *Statistics Through Applications*, First and Second Editions.

JOSH TABOR has enjoyed teaching general and AP® Statistics to high school students for more than 18 years, most recently at his alma mater, Canyon del Oro High School in Oro Valley, Arizona. He received a BS in Mathematics from Biola University, in La Mirada, California. In recognition of his outstanding work as an educator, Josh was named one of the five finalists for Arizona Teacher of the Year in 2011. He is a past member of the AP® Statistics Development Committee (2005–2009), as well as an experienced Table Leader and Question Leader at the AP® Statistics Reading. Each year, Josh leads one-week AP® Summer Institutes and one-day College Board workshops around the country and frequently speaks at local, national, and international conferences. In addition to teaching and speaking, Josh has authored articles in *The Mathematics Teacher*, *STATS Magazine*, and *The Journal of Statistics Education*. He is the author of the *Annotated Teacher's Edition* and *Teacher's Resource Materials* for *The Practice of Statistics* 4e and 5e, along with the *Solutions Manual* for *The Practice of Statistics* 5e. Combining his love of statistics and love of sports, Josh teamed with Christine Franklin to write *Statistical Reasoning in Sports*, an innovative textbook for on-level statistics courses.

DANIEL S. YATES taught AP® Statistics in the Electronic Classroom (a distance-learning facility) affiliated with Henrico County Public Schools in Richmond, Virginia. Prior to high school teaching, he was on the mathematics faculty at Virginia Tech and Randolph-Macon College. He has a PhD in Mathematics Education from Florida State University. Dan received a College Board/Siemens Foundation Advanced Placement Teaching Award in 2000.

DAVID S. MOORE is Shanti S. Gupta Distinguished Professor of Statistics (Emeritus) at Purdue University and was 1998 President of the American Statistical Association. David is an elected fellow of the American Statistical Association and of the Institute of Mathematical Statistics and an elected member of the International Statistical Institute. He has served as program director for statistics and probability at the National Science Foundation. He is the author of influential articles on statistics education and of several leading textbooks.

Content Advisory Board and Supplements Team

Jason Molesky, Lakeville Area Public School, Lakeville, MN
Media Coordinator—Worked Examples, PPT lectures, Strive for a 5 Guide

Jason has served as an AP® Statistics Reader and Table Leader since 2006. After teaching AP® Statistics for eight years and developing the FRAPPY system for AP® exam preparation, Jason moved into administration. He now serves as the Director of Program Evaluation and Accountability, overseeing the district's research and evaluation, continuous improvement efforts, and assessment programs. Jason also provides professional development to statistics teachers across the United States and maintains the "Stats Monkey" Web site, a clearinghouse for AP® Statistics resources.

Tim Brown, The Lawrenceville School, Lawrenceville, NJ
Content Advisor, Test Bank, TRM Tests and Quizzes

Tim first piloted an AP® Statistics course the year before the first exam was administered. He has been an AP® Reader since 1997 and a Table Leader since 2004. He has taught math and statistics at The Lawrenceville School since 1982 and currently holds the Bruce McClellan Distinguished Teaching Chair.

Doug Tyson, Central York High School, York, PA
Exercise Videos, Learning Curve

Doug has taught mathematics and statistics to high school and undergraduate students for 22 years. He has taught AP® Statistics for 7 years and served as an AP® Reader for 4 years. Doug is the co-author of a curriculum module for the College Board, conducts student review sessions around the country, and gives workshops on teaching statistics.

Paul Buckley, Gonzaga College High School, Washington, DC
Exercise Videos

Paul has taught high school math for 20 years and AP® Statistics for 12 years. He has been an AP® Statistics Reader for six years and helps to coordinate the integration of new Readers (Acorns) into the Reading process. Paul has presented at Conferences for AP®, NCTM, NCEA (National Catholic Education Association) and JSEA (Jesuit Secondary Education Association).

Leigh Nataro, Moravian Academy, Bethlehem, PA
Technology Corner Videos

Leigh has taught AP® Statistics for nine years and has served as an AP® Statistics Reader for the past four years. She enjoys the challenge of writing multiple-choice questions for the College Board for use on the AP® Statistics exam. Leigh is a National Board Certified Teacher in Adolescence and Young Adulthood Mathematics and was previously named a finalist for the Presidential Award for Excellence in Mathematics and Science Teaching in New Jersey.

Ann Cannon, Cornell College, Mount Vernon, IA
Content Advisor, Accuracy Checker

Ann has served as Reader, Table Leader, and Question Leader for the AP® Statistics exam for the past 13 years. She has taught introductory statistics at the college level for 20 years and is very active in the Statistics Education Section of the American Statistical Association, serving on the Executive Committee for two 3-year terms. She is co-author of *STAT2: Building Models for a World of Data* (W. H. Freeman and Company).

Michel Legacy, Greenhill School, Dallas, TX
Content Advisor, Strive for a 5 Guide

Michael is a past member of the AP® Statistics Development Committee (2001–2005) and a former Table Leader at the Reading. He currently reads the Alternate Exam and is a lead teacher at many AP® Summer Institutes. Michael is the author of the 2007 College Board AP® Statistics Teacher's Guide and was named the Texas 2009–2010 AP® Math/Science Teacher of the Year by the Siemens Corporation.

James Bush, Waynesburg University, Waynesburg, PA
Learning Curve, Media Reviewer

James has taught introductory and advanced courses in Statistics for over 25 years. He is currently a Professor of Mathematics at Waynesburg University and is the recipient of the Lucas Hathaway Teaching Excellence Award. James has served as an AP® Statistics Reader for the past seven years and conducts many AP® Statistics preparation workshops.

Beth Benzing, Strath Haven High School, Wallingford/Swarthmore School District, Wallingford, PA
Activities Videos

Beth has taught AP® Statistics for 14 years and has served as a Reader for the AP® Statistics exam for the past four years. She serves as Vice President on the board for the regional affiliate for NCTM in the Philadelphia area and is a moderator for an on-line course, *Teaching Statistics with Fathom*.

Heather Overstreet, Franklin County High School, Rocky Mount, VA
TI-Nspire Technology Corners

Heather has taught AP® Statistics for nine years and has served as an AP® Statistics Reader for the past six years. While working with Virginia Advanced Study Strategies, a program for promoting AP® math, science, and English courses in Virginia High Schools, she led many AP® Statistics Review Sessions and served as a Laying the Foundation trainer of teachers of pre-AP® math classes.

Acknowledgments

Teamwork: It has been the secret to the successful evolution of *The Practice of Statistics* (*TPS*) through its fourth and fifth editions. We are indebted to each and every member of the team for the time, energy, and passion that they have invested in making our collective vision for *TPS* a reality.

To our team captain, Ann Heath, we offer our utmost gratitude. Managing a revision project of this scope is a Herculean task! Ann has a knack for troubleshooting thorny issues with an uncanny blend of forthrightness and finesse. She assembled an all-star cast to collaborate on the fifth edition of *TPS* and trusted each of us to deliver an excellent finished product. We hope you'll agree that the results speak for themselves. Thank you, Ann, for your unwavering support, patience, and friendship throughout the production of these past two editions.

Truth be told, we had some initial reservations about adding a Development Editor to our team starting with the fourth edition of *TPS*. Don Gecewicz quickly erased our doubts. His keen mind and sharp eye were evident in the many probing questions and insightful suggestions he offered at all stages of the project. Thanks, Don, for your willingness to push the boundaries of our thinking.

Working behind the scenes, Enrico Bruno busily prepared the manuscript chapters for production. He did yeoman's work in ensuring that the changes we intended were made as planned. For the countless hours that Enrico invested sweating the small stuff, we offer him our sincere appreciation.

We are deeply grateful to Patrice Sheridan and to Diana Blume for their aesthetic contributions to the eye-catching design of *TPS* 5e. Our heartfelt thanks also go to Vivien Weiss and Susan Wein at W. H. Freeman for their skillful oversight of the production process. Patti Brecht did a superb job copyediting a very complex manuscript.

A special thank you goes to our good friends on the high school sales and marketing staff at Bedford, Freeman, and Worth (BFW) Publishers. We feel blessed to have such enthusiastic professionals on our extended team. In particular, we want to thank our chief cheerleader, Cindi Weiss, for her willingness to promote *The Practice of Statistics* at every opportunity.

We cannot say enough about the members of our Content Advisory Board and Supplements Team. This remarkable group is a veritable who's who of the AP® Statistics community. We'll start with Ann Cannon, who once again reviewed the statistical content of every chapter. Ann also checked the solutions to every exercise in the book. What a task! More than that, Ann offered us sage advice about virtually everything between the covers of *TPS* 5e. We are so grateful to Ann for all that she has done to enhance the quality of *The Practice of Statistics* over these past two editions.

Jason Molesky, aka "Stats Monkey," greatly expanded his involvement in the fifth edition by becoming our Media Coordinator in addition to his ongoing roles as author of the *Strive for a 5 Guide* and creator of the PowerPoint presentations for the book. Jason also graciously loaned us his Free Response AP® Problem, Yay! (FRAPPY!) concept for use in *TPS* 5e. We feel incredibly fortunate to have such a creative, energetic, and deeply thoughtful person at the helm as the media side of our project explodes in many new directions at once.

Jason is surrounded by a talented media team. Doug Tyson and Paul Buckley have expertly recorded the worked exercise videos. Leigh Nataro has produced screencasts for the Technology Corners on the TI-83/84, TI-89, and TI-Nspire. Beth Benzing followed through on her creative suggestion to produce "how to" screencasts for teachers to accompany the Activities in the book. James Bush capably served as our expert reviewer for all of the media elements and is now partnering with Doug Tyson on a new Learning Curve media component. Heather Overstreet once

again compiled the TI-Nspire Technology Corners in Appendix B. We wish to thank the entire media team for their many contributions.

Tim Brown, my Lawrenceville colleague, has been busy creating many new, high-quality assessment items for the fifth edition. The fruits of his labors are contained in the revised *Test Bank* and in the quizzes and tests that are part of the *Teacher's Resource Materials*. We are especially thankful that Tim was willing to compose an additional cumulative test for Chapters 2 through 12.

We offer our continued thanks to Michael Legacy for composing the superb questions in the Cumulative AP® Practice Tests and the *Strive for a 5 Guide*. Michael's expertise as a two-time former member of the AP® Statistics Test Development Committee is invaluable to us.

Although Dan Yates and David Moore both retired several years ago, their influence lives on in *TPS* 5e. They both had a dramatic impact on our thinking about how statistics is best taught and learned through their pioneering work as textbook authors. Without their early efforts, *The Practice of Statistics* would not exist!

Thanks to all of you who reviewed chapters of the fourth and fifth editions of *TPS* and offered your constructive suggestions. The book is better as a result of your input.

—Daren Starnes and Josh Tabor

A *final note from Daren*: It has been a privilege for me to work so closely with Josh Tabor on *TPS* 5e. He is a gifted statistics educator; a successful author in his own right; and a caring parent, colleague, and friend. Josh's influence is present on virtually every page of the book. He also took on the thankless task of revising all of the solutions for the fifth edition in addition to updating his exceptional *Annotated Teacher's Edition*. I don't know how he finds the time and energy to do all that he does!

The most vital member of the *TPS* 5e team for me is my wonderful wife, Judy. She has read page proofs, typed in data sets, and endured countless statistical conversations in the car, in airports, on planes, in restaurants, and on our frequent strolls. If writing is a labor of love, then I am truly blessed to share my labor with the person I love more than anyone else in the world. Judy, thank you so much for making the seemingly impossible become reality time and time again. And to our three sons, Simon, Nick, and Ben—thanks for the inspiration, love, and support that you provide even in the toughest of times.

A *final note from Josh*: When Daren asked me to join the *TPS* team for the fourth edition, I didn't know what I was getting into. Having now completed another full cycle with *TPS* 5e, I couldn't have imagined the challenge of producing a textbook—from the initial brainstorming sessions to the final edits on a wide array of supplementary materials. In all honesty, I still don't know the full story. For taking on all sorts of additional tasks—managing the word-by-word revisions to the text, reviewing copyedits and page proofs, encouraging his co-author, and overseeing just about everything—I owe my sincere thanks to Daren. He has been a great colleague and friend throughout this process.

I especially want to thank the two most important people in my life. To my wife Anne, your patience while I spent countless hours working on this project is greatly appreciated. I couldn't have survived without your consistent support and encouragement. To my daughter Jordan, I look forward to being home more often and spending less time on my computer when I am there. I also look forward to when you get to use *TPS* 7e in about 10 years. For now, we have a lot of fun and games to catch up on. I love you both very much.

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.

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.

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

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.

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

HOW TO MAKE A SCATTERPLOT

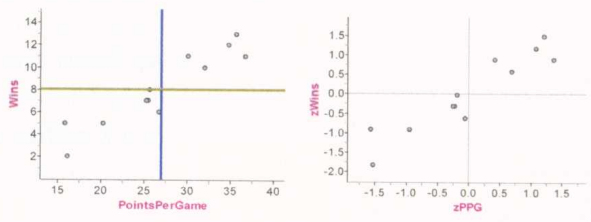
1. Decide which variable should go on each axis.
2. Label and scale your axes.
3. Plot individual data values.

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

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 or lower-left "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.

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

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 $\hat{y} = b_0 + b_1x$. We prefer our simpler versions without the subscripts!

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 o having the same color (say, red). The parameter is tl chips in the population: $p = 0.50$. In this Activity, y 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 sampl When finished, the student should return all the chi and pass the bag to the next student.

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

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?

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.

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 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.

CHECK YOUR UNDERSTANDING

Identify the explanatory and response variables in each setting.

1. 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.

2. 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.

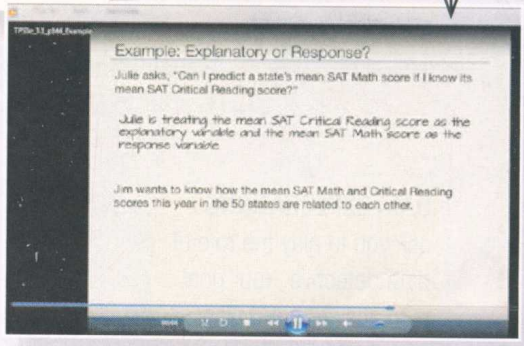
PROBLEM: For each student, identify the explanatory variable and the response variable if possible.

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.

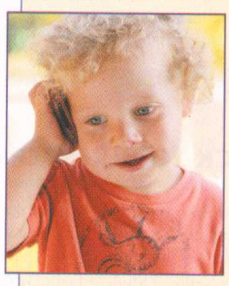
1. pg 144 **Coral reefs** How sensitive to changes in water temperature are coral reefs? To find out, measure the growth of corals in aquariums where the water temperature is controlled at different levels. Growth is measured by weighing the coral before and after the experiment. What are the explanatory 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.

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.

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 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?

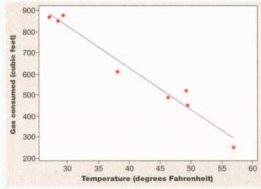
36. **What's my line?** An eccentric professor believes that a child with IQ 100 should have a reading test score of 50 and predicts that reading score should increase by 1 point for every additional point of IQ. What is the equation of the professor's regression line for predicting reading score from 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 (city mpg).
(a) What's the slope of this line? Interpret this value in context.
(b) What's the y intercept? Explain why the value of the intercept is not statistically meaningful.
(c) 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(IQ sco
(a) What's the slope of this line? Interpret thi context.
(b) What's the y intercept? Explain why the intercept is not statistically meaningful.
(c) Find the predicted reading score for a cl IQ score of 90.

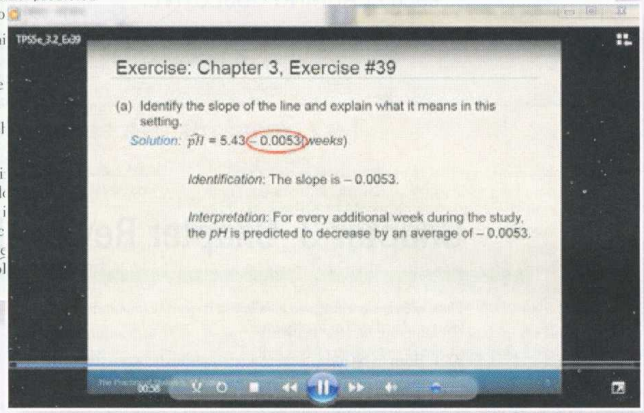
39. **Acid rain** Researchers studying acid rai pg 166 the acidity of precipitation in a Colorad area for 150 consecutive weeks. Acidity i by pH. Lower pH values show higher ac researchers observed a linear pattern ov They reported that the regression line $\hat{pH}$ = 5.43 − 0.0053(weeks) fit the data well.[19]

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 $\hat{y} = 1425 - 19.87x$.



(a) Identify the slope of the line and explain what it means in this setting.
(b) Identify the y intercept of the line. Explain why it's risky to use this value as a prediction.
(c) 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.



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 $r = 1$ indic ing a perfectly linear relationship. Strong linear assoc tions have correlations near 1 or -1, while weak lin 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.

What Did You Learn?

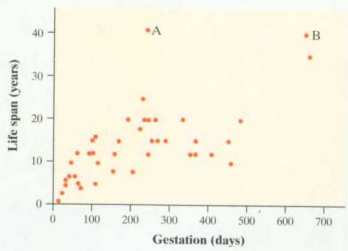
Learning Objective	Section	Related Example on Page(s)	Relevant Chapter Review Exercise(s)
Identify explanatory and response variables in situations where one variable helps to explain or influences the other.	3.1	144	R3.4
Make a scatterplot to display the relationship between two quantitative variables.	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. Explain the dangers of extrapolation.	3.2	167, Discussion on 168 (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

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

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]

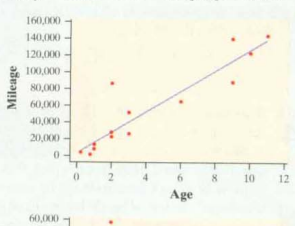


(a) Describe the association shown in the scatterplot.

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.