

# STATISTICS

THE ART AND SCIENCE OF LEARNING FROM DATA



*Agresti • Franklin*



# Statistics

## The Art and Science of Learning from Data

**Alan Agresti**  
University of Florida

**Christine Franklin**  
University of Georgia



Upper Saddle River, New Jersey 07458

**Library of Congress Cataloging-in-Publication Data**

Agresti, Alan.

Statistics: the art and science of learning from data/Alan Agresti, Christine A. Franklin.

p. cm.

ISBN 0-13-008369-0—ISBN 0-13-045536-9 (IE)

1. Statistics—Textbooks. I. Franklin, Christine A. II. Title

QA276.12.A35 2007

519.5—dc22

2005048929

Executive Acquisitions Editor: *Petra Recter*

Editor in Chief: *Sally Yagan*

Executive Project Manager: *Ann Heath*

Project Manager: *Michael Bell*

Production Editor: *Lynn Savino Wendel*

Senior Managing Editor: *Linda Mihatov Behrens*

Executive Managing Editor: *Kathleen Schiapparelli*

Manufacturing Manager: *Alexis Heydt-Long*

Manufacturing Buyer: *Alan Fischer*

Director of Marketing: *Patrice Jones*

Managing Editor, Digital Supplements: *Nicole M. Jackson*

Media Production Editor: *John Cassar*

Development Editor: *Donald Gecewicz*

Editorial Assistant / Supplements Editor: *Joanne Wendelken*

Art Director: *Maureen Eide*

Assistant to Art Director: *Dina Curro*

Interior Designer: *Susan Anderson*

Cover Designer: *Maureen Eide / Suzanne Behnke*

Art Editor: *Thomas Benfatti*

Creative Director: *Juan R. López*

Director of Creative Services: *Paul Belfanti*

Director, Image Resource Center: *Melinda Reo*

Manager, Rights and Permissions: *Zina Arabia*

Manager, Visual Research: *Beth Brenzel*

Manager, Cover Visual Research and Permissions: *Karen Sanatar*

Image Permission Coordinator: *Richard Rodrigues*

Photo Researcher: *Sheila Norman*

Manager, Composition: *Allyson Graesser*

Assistant Manager, Composition: *Clara Bartunek*

Electronic Production Specialist: *Julita Nazario*

Art Studio: Precision Graphics, Inc. / Laserwords

Cover Photo Credits: Jon Shireman / Stone / Getty Images, Inc.



© 2007 by Pearson Education, Inc.

Pearson Prentice Hall

Pearson Education, Inc.,

Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.

Printed in the United States of America

10 9 8 7 6 5 4 3 2

ISBN 0-13-008369-0

Pearson Education Ltd., London

Pearson Education Australia Pty. Limited, Sydney

Pearson Education Singapore Pte. Ltd.

Pearson Education North Asia, Ltd, Hong Kong

Pearson Education Canada, Ltd., Toronto

Pearson Educación de México, S.A., de C.V.

Pearson Education, Japan, Tokyo

Pearson Education Malaysia, Pte. Ltd.

## *Dedication*

*To my wife Jacki for her extraordinary support, including making numerous suggestions and putting up with the evenings and weekends I was working on this book, and to my good friend Bruce Edwards for his continual encouragement over four years of weekly lunches.*

*Alan Agresti*

*To Corey and Cody, who have shown me the joys of motherhood and to my husband, Dale, for being a dear friend and a dedicated father to our boys.*

*Chris Franklin*



## Agresti/Franklin CD Contents

### Instructor to Instructor videos (on the Instructor Edition CD only)

#### Data Sets:

1. MINITAB
2. TI-83 Series, TI-84 Series, TI-89 Series
3. Excel
4. SPSS
5. JMP® INTRO
6. ASCII

Look at the back endpapers of the book for a complete list of data files. Information is given on which examples, exercises, and activities require or reference the data files and applets.

#### Applets:

1. Sample from a population
2. Sampling distributions
3. Random numbers
4. Long run probability demonstrations:
  - a. Simulating the probability of rolling a 6
  - b. Simulating the probability of rolling a 3 or 4
  - c. Simulating the probability of head with a fair coin
  - d. Simulating the probability of head with an unfair coin ( $P(H) = 0.2$ )
  - e. Simulating the probability of head with an unfair coin ( $P(H) = 0.8$ )
  - f. Simulating the stock market
5. Mean versus median
6. Standard deviation
7. Confidence intervals for a proportion
8. Confidence intervals for a mean (the impact of confidence level)
9. Hypothesis tests for a proportion
10. Hypothesis tests for a mean
11. Correlation by eye
12. Regression by eye

**Activities Manual** — containing additional activities to those found within the text.

# Preface

*“Statistics . . . the most important science in the whole world: for upon it depends the practical application of every other science and of every art; the one science essential to all political and social administration, all education, all organization based upon experience, for it only gives the results of our experience.”*

FLORENCE NIGHTINGALE (1820–1910)

One of Florence Nightingale’s achievements as a nurse was to promote improved health care by using the art and science of statistics to help people learn from data. She was one of the individuals who helped to develop a new way of looking at the world that is now vital in this modern information age. Today, everyone from students and teachers to workers and employers has to be able to interpret data.

## Why Did We Write This Book?

We have each taught introductory statistics for more than 25 years, and have witnessed the welcome evolution from the traditional formula-driven watered-down mathematical statistics course to a concept-driven approach. This concept-driven approach places more emphasis on why statistics is important in the real world and places less emphasis on probability. Our book’s content and approach reflect our philosophy of teaching, our classroom experience, recent advances in technology, and student feedback about how they believe that they learn most easily. One of our main purposes in writing this book was to help make the conceptual approach more interesting and more readily accessible to the average college student. At the end of the course, we want more students to look back at their statistics course and think that they learned many worthwhile ideas that will serve them well for the rest of their lives.

## What Approach Does This Book Take?

In 2005, the American Statistical Association (ASA) endorsed guidelines and recommendations for the introductory statistics course as described in the report, “Guidelines for Assessment and Instruction in Statistics Education (GAISE) for the College Introductory Course” ([www.amstat.org/education/gaise](http://www.amstat.org/education/gaise)). The report states that the overarching goal of all introductory statistics courses is to produce statistically educated students which means that students should develop statistical literacy and the ability to think statistically. The report gives six key recommendations for the college introductory course:

- Emphasize statistical literacy and develop statistical thinking.
- Use real data.
- Stress conceptual understanding rather than mere knowledge of procedures.
- Foster active learning in the classroom.
- Use technology for developing concepts and analyzing data.
- Use assessment to evaluate and improve student learning.



We wholeheartedly endorse these recommendations and have endeavored to write a textbook that supports these guidelines.

Here are the main ways that *Statistics: The Art and Science of Learning from Data* attempts to provide an improved context for teaching introductory statistics:

**Learn to ask and answer interesting questions:**

In presenting concepts and methods, we encourage students to think about the data and the appropriate analyses by continually posing questions. Our approach, learning by framing questions, is carried out in various ways, including (1) presenting a structured approach to examples that separates the question and the analysis from the scenario presented, (2) providing lots of homework problems that encourage students to think and write, (3) requesting that students perform simulations and hands-on activities to help them learn by doing, and (4) posing information in the form of questions in the section titles and in the figure captions that are answered within the text.

**Keep students' needs in mind and reduce confusion:**

In many class tests, our students have told us that this book is more “readable” than other introductory statistics texts. To achieve this readability, we have included a wide array of interesting real-data examples and exercises that intrigue students. We have also attempted to simplify our prose wherever possible, without sacrificing any of the accuracy that instructors expect in a textbook.

A serious source of confusion for students is the multiplicity of inference methods that derive from the many combinations of confidence intervals and tests, means and proportions, large sample and small sample, variance known and unknown, two-sided and one-sided inference, independent and dependent samples, and so on. We’ve placed the main emphasis on the most important cases for practical application of inference: large sample, variance unknown, two-sided inference, and independent samples. The many other cases are also covered (except for known variances), but more briefly, with the exercises focusing mainly on the way inference is commonly conducted in practice. We believe that this approach will help to reduce students’ confusion.

**Connect statistics to the real world through a more balanced presentation of means and proportions:**

Every day in the media, we see and hear percentages and rates being used to summarize results of opinion polls, outcomes of medical studies, and economic reports. Less often do we hear about outcomes that are in any sense normally distributed or even continuous, yet that’s the main emphasis of most statistics texts. In this text, we have increased the attention paid to the analysis of proportions. For example, we use contingency tables early in the text to illustrate the concept of association between two categorical variables and to show the potential influence of a lurking variable. When introducing sampling distributions and inference methods, we first explain using proportions for a binary outcome.

We don’t neglect quantitative data, but we believe that students should become comfortable with data in the form that they will most often see in practice. This approach has pedagogical rewards. We have found that it is easier for a student to look at a sample of binary data or a distribution of sample proportions with a small sample size and understand what that distribution refers to than to do the same for a bell-shaped curve for a continuum of unobserved values.

## How Have We Organized the Contents?

We've shaped the organization and content based on extensive feedback from reviewers and class testers and our beliefs about what works in the classroom and which topics merit inclusion in an introductory course. The main focus is on concepts and interpretation rather than on blind plugging of numbers into formulas. We also want students to come to appreciate that in practice, assumptions are not perfectly satisfied, models are not exactly correct, distributions are not exactly normal, and all sorts of factors should be considered in conducting a statistical analysis. The title of our book reflects the experience of data analysts, who soon realize that statistics is an art as well as a science.

Here's a brief chapter-by-chapter survey. We summarize the content of each chapter and also point out ways in which our presentation may differ somewhat from other texts:

### **Chapter 1: Statistics: The Art and Science of Learning from Data**

Our first chapter gives an overview to students about how data and statistical analyses surround us in our everyday lives by presenting informative examples of real studies. The General Social Survey (GSS) is introduced, and instructions are given for downloading useful and interesting data that will be used throughout the textbook. We also introduce the use of technology through software and applets. An introductory overview of the field such as this chapter provides is missing from many statistics texts.

### **Chapter 2: Exploring Data with Graphs and Numerical Summaries**

This first of two chapters on descriptive statistics introduces students to the univariate case—graphical and numerical methods for exploring and summarizing data on a single categorical or quantitative variable. The emphasis is on interpreting graphs and numerical summaries produced by computer or calculator output, rather than on constructing graphs and calculating numerical summaries by hand.

### **Chapter 3: Association: Contingency, Correlation, and Regression**

This second of two chapters on descriptive statistics focuses on bivariate analyses. It introduces one of the most important topics of statistics—analyzing association between two variables. The chapter begins by exploring categorical variables through contingency tables. Students often see such summaries in their everyday lives, yet most books delay coverage of this material until near the end. Discussing contingency tables facilitates transition into Chapter 4 on producing data for topics such as survey sampling and comparisons in randomized experiments. It also provides background for the discussion of conditional probability in Chapter 5. We next introduce correlation and regression methods for quantitative variables. For many students, the first time that they see correlation is a “light-bulb” moment, the instant they understand the power and usefulness of statistics. The final section emphasizes important cautions, such as association not proving causation and the potential influence of lurking variables, two fundamental concepts that students must see early in the course.

### **Chapter 4: Gathering Data**

This chapter introduces students to good ways and poor ways to sample and to experiment. Why an entire chapter on producing data? Valid analysis and interpretation of data depend on having an appropriate study design. When is it appropriate to make conclusions about causes and effects? How could other factors influence results? How important are issues such as nonresponse? What



determines the extent to which conclusions can be generalized? Building on the presentation from Chapter 3 with proportions, we informally introduce the concept of margin of error—a concept and term seen daily by students reading about sample surveys in the media. Regardless of how sophisticated a statistical analysis is conducted, results may be nearly worthless if the study design is poor. However, even though randomized experiments are ideal, they are often not feasible or ethical. In practice, the types of studies we most often encounter are observational. Thus, we also devote detailed coverage to types of observational studies, with emphasis on sample surveys. The potential influence of lurking variables is further illustrated with examples of observational studies.

## **Chapter 5: Probability in Our Daily Lives**

This chapter introduces students to the basic idea of probability and to how it can help us understand the forms that randomness takes in our lives. The unifying theme is how probability relates to our daily lives, using applications such as diagnostic disease testing and whether or not an apparent coincidence truly is coincidental. It's easy for students to get lost when probability is a major component of the course or is presented in a mathematical manner. So we attempt to offer “just enough” to help students understand probability distributions and sampling distributions in the following chapter and conditional probability in the analyses of contingency tables. The early introduction of contingency tables and conditional proportions in Chapter 3 lays the foundation for students to understand more easily the difficult concept of conditional probability. Students are also shown that simulation can estimate probabilities not easily calculated by simple rules.

## **Chapter 6: Probability Distributions**

This chapter focuses on two of the most important probability distributions—the normal for continuous random variables and the binomial for discrete random variables—and introduces the fundamental concept of sampling distributions. The sampling distribution is a topic that students find hard to grasp, yet understanding it is crucial to their understanding of statistical inference. We believe sampling distributions are simplest to explain first for categorical data—using proportions in a small-sample example that identifies easily all the possible samples of binary data. The sampling distribution for a sample proportion is developed using a natural progression moving from the binomial distribution of counts. We believe that students who work through these sections and also the applet-related activities will have a stronger understanding of sampling distributions than students usually achieve in introductory-statistics courses. The end of the chapter links sampling distributions to the basic ideas of statistical inference, which is the focus of the following two chapters.

## **Chapter 7: Statistical Inference: Confidence Intervals**

We believe that more emphasis should be placed on interval estimation than on significance testing in an introductory statistics course. We learn less from testing that a parameter takes some specific value than from finding a range of plausible values for it. Statisticians increasingly share the belief that significance tests have been overemphasized. For instance, many psychology journals now forbid reporting P-values without accompanying information about effect sizes, such as confidence intervals. Moreover, confidence intervals are simpler than tests for students to understand, especially with the use of an applet for simulation. Consequently, we devote this first inference chapter solely to estimation. An activity using simulation helps students visualize the role of the sampling distribution in forming confidence intervals. We begin the chapter with forming a confidence interval for a proportion. We believe that this is an easier transition for the students, because this method uses the normal distribution. We then move to the confidence interval for a mean, introducing the  $t$  distribution. We do not cover the case in which the population

standard deviation is known, as this is very rarely true in practice and increases the profusion of methods that leads to student confusion.

## **Chapter 8: Statistical Inference: Significance Tests About Hypotheses**

Although we believe it's important to emphasize interval estimation for inference, significance tests do play an important role in statistical inference. The main goal in many research studies is to check whether the data support certain statements, claims, or predictions. A significance test is a method of using data to summarize evidence about a claim or prediction. This chapter emphasizes the P-value approach, which is the common way to perform tests in practice and is more informative than simply making a reject-or-not-reject decision at an artificial alpha-level with a rejection region. Most journals prefer using P-values to report statistical results. As in the previous two chapters, we begin by introducing the significance test for a proportion, then move to the significance test for a mean. Both one-sided and two-sided alternatives are presented. We encourage more use of the two-sided alternative, however, because this is commonly seen in practice and because the results of using a two-sided alternative parallel (two-sided) confidence intervals. Also, in that case,  $t$  methods are robust to violations of the normality assumption. In this chapter, we also stress the limitations of significance tests. In doing so, we emphasize the distinction between *practical significance* and *statistical significance*, and we show the possible difficulties resulting when only "significant" results are reported. We also discuss Type I error, Type II error, and the power of a test.

## **Chapter 9: Comparing Two Groups**

The chapter explains that a comparison of two groups is actually a bivariate analysis, in which the explanatory variable defines the groups to be compared. We avoid the case of known population standard deviation and we de-emphasize the analysis that assumes equal variances, showing how software can find the  $df$  value with the less restrictive analysis. We present methods for independent and for dependent samples (matched pairs), with our main emphasis on the independent samples. Because this text does not short-change categorical data, we also present methods for analyzing proportions using matched pairs. This topic is relatively simple to present but is rarely treated in introductory texts. This chapter also revisits the importance of taking into account potential lurking variables when comparing two groups, illustrating with an example showing Simpson's paradox.

## **Chapter 10: Analyzing the Association Between Categorical Variables**

In our opinion, introductory statistics books place way too much emphasis on the chi-squared test. Like any significance test, it gives limited information. It should not be the sole or even the primary method taught in the inference chapter about bivariate categorical data. The chi-squared test addresses whether an association exists between categorical variables; however, it does not address the strength or the pattern of the association. We have included follow-up analyses such as measuring association using the difference and the ratio of proportions (relative risk), and inspecting residuals to detect which cells are responsible for a large chi-squared statistic. Doing so connects this material with regression methods presented in the following chapter for bivariate quantitative data.

## **Chapter 11: Analyzing Association Between Quantitative Variables: Regression Analysis**

This chapter expands on the concepts of simple straight-line regression introduced in Chapter 3. After reviewing the introductory material from Chapter 3 on straight-line regression equations and correlation, this chapter explains



what is meant by a model, connects this with the notion of residual unexplained variation around a regression line (and the related ANOVA table and  $r$ -squared measure and  $F$  statistic), discusses “regression toward the mean,” and presents inference methods. The chapter also discusses ways of checking model assumptions.

### Chapter 12–14: Advanced Topics

Chapters 12–14 contain more advanced topics for the first course or for a follow-up course. Chapter 12 covers multiple regression. It also shows how a regression model can include categorical explanatory variables (using indicator variables) or categorical response variables (logistic regression). Chapter 13 covers one-way and two-way ANOVA and shows their connection with multiple regression. Chapter 14 introduces nonparametric statistics.

## How Do We Promote Student Learning?

In an effort to motivate students to think about the material, to ask appropriate questions, and to develop good problem-solving skills, we have created special features that distinguish this text.

### A Vivid Chapter-Opening Example

Each chapter begins with a high-interest example that establishes some key questions and themes that are woven throughout the chapter. Illustrated with engaging photographs, this example is designed to ‘grab students’ attention and to draw them into the chapter. The issues discussed in the chapter’s opening example are referred to and revisited in examples within the chapter.

### Our Innovative Example Format

Recognizing that the worked examples are the key vehicle for engaging and teaching students, we have developed a unique structure to help students learn to model the question-posing and investigative thought process required to examine issues intelligently using statistics. The five components in this structure are as follow:

- **Picture the Scenario:** Background information is presented so students can visualize the situation. This step places the data to be investigated in context and often provides a link back to previous examples.
- **Questions to Explore:** Questions are posed that follow directly from information in the scenario. This step helps the student to focus on what is to be learned from the example and what types of questions are useful to ask.
- **Think It Through:** This step is the heart of each example as the questions posed are investigated and answered using appropriate statistical methods. Each solution is clearly matched to the questions so students can easily find the response to each posed Question to Explore.
- **Insight:** The overarching goal of the Insight step is to clarify the central ideas investigated in the example and to place them in a broader context. Many of the Insights also provide connections between seemingly disparate topics in the text by referring back to concepts learned previously and/or foreshadowing techniques and ideas to come later.
- **To Practice This Concept, Try Exercise XX:** Each example concludes by directing students to an end of section exercise that allows immediate practice of the concept or technique within the example, requiring the student to use similar reasoning as that employed in the example.

This example, from Chapter 2, demonstrates our example format:

### EXAMPLE 9

## Is There a Trend Toward Warming in New York City?



**Picture the Scenario:** Background information is presented so students can visualize the situation. This step places the data to be investigated in context and often provides a link back to previous examples.

**Questions to Explore:** Questions are posed that follow directly from information in the scenario. Posing questions helps the student to focus on what is to be learned from the example and what types of questions are useful to ask about the data.

**Think It Through:** This step is the heart of each example as the questions posed are investigated and answered using appropriate statistical methods. Each solution is clearly matched to the question so students can easily find the response to each posed Question to Explore.

### Picture the Scenario

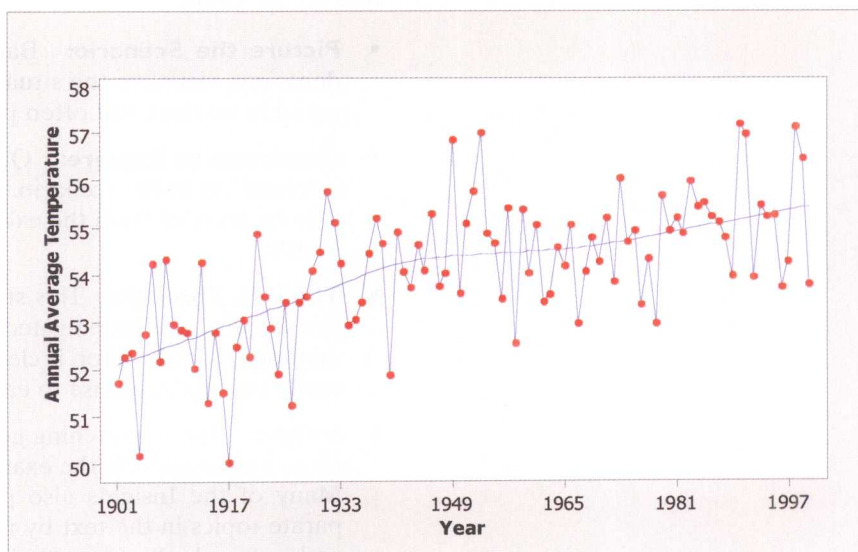
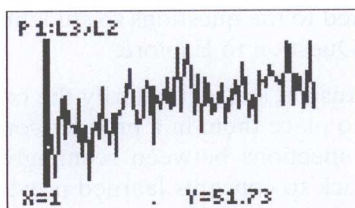
Each day, weather stations around the world record the daily average temperature, defined as the average of the maximum and minimum temperature over a midnight-to-midnight time frame. In a given year, the annual average temperature is the average of the daily average temperatures for that year. Let's analyze data on annual average temperature (in degrees Fahrenheit) in Central Park, New York City from 1901 to 2000. This is a continuous quantitative variable. The data are in the "Central Park Yearly Temps" data file on the text CD.

### Question to Explore

What do we learn from a time plot of these annual average temperatures? Is there a trend toward warming in the United States?

### Think It Through

Figure 2.8 shows a time plot, constructed using MINITAB software. The time plot constructed using the TI-83-84 calculator is in the margin. The observations fluctuate considerably, but the figure does suggest somewhat of an increasing trend in the annual average temperatures at this location.



▲ **FIGURE 2.8** MINITAB output for a time plot of Central Park, New York City average annual temperatures. The annual average temperatures are plotted against the year from 1901 to 2000. A smoothing curve is superimposed. **Question:** Are the annual average temperatures tending to increase, decrease, or stay the same over time?



**Insight:** This step clarifies the central ideas investigated in the example and places them in a broader context that often states the conclusions in less technical terms. Many of the Insights also provide connections between seemingly disparate topics in the text by referring back to concepts learned previously and /or foreshadowing techniques and ideas to come.

**To Practice This Concept:** Each example concludes by directing students to an end-of-section exercise that allows immediate practice of the concept or technique within the example.

### Insight


The short-term fluctuations in a time plot can mask an overall trend. It's possible to get a clearer picture by "smoothing" the data. This is beyond our scope here. MINITAB presents the option of smoothing the data to portray a general trend (Click on "data view" and choose the "lowess" option under "smoother"). This is the smooth curve passing through the data points in Figure 2.8. This curve goes from a level of about 52 degrees in 1901 to about 55 degrees in 2000.

The data reported in Figure 2.8 refer to one location in the United States. In a study of climate change, it would be important to explore other locations around the world to see if similar trends are evident there.

**TO PRACTICE THIS CONCEPT, TRY EXERCISES 2.26 AND 2.27**

## A Broad Array of Useful and Engaging Exercises

The text contains a strong emphasis on real data both in the examples and exercises. Nearly all the chapters contain more than 100 exercises. These exercises are realistic and ask students to provide interpretations of the data or scenario rather than merely to find a numerical solution. We show how statistics addresses a wide array of applications, from opinion polls, to market research, to health and human behavior. Our goal is to have students come to believe the comment by the author H. G. Wells (1866–1946), "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

Exercises are placed at the end of each section for immediate reinforcement, and a more comprehensive set of exercises is found at the end of the chapter. Each exercise has a descriptive label. Exercises for which technology is recommended are indicated with the icon, . Larger data sets used in examples and exercises are referenced in text, listed in the back endpapers and made available on the text CD. The exercises are divided into the following categories:

- **Practicing the Basics:** The section exercises and the first group of end-of-chapter exercises reinforce basic application of the methods.
- **Concepts and Investigations:** These exercises require that the student explore real data sets and carry out investigations for mini-projects, and explore concepts and related theory or extensions of the chapter's methods. This section contains some multiple-choice and true/false exercises to help students check their understanding of the basic concepts and to prepare for tests. It also contains a few more difficult, optional exercises (highlighted with the ♦♦ icon) that present some additional concepts and methods.
- **Class Explorations:** A brief third section provides exercises that are designed for group work based on investigations done by each of the students in a team. Additional activities may be found here as well.

## Applying Statistics to Real Situations: In Practice

One of our goals is to help students understand some of the differences between proper "academic" statistics and what is actually done in practice. Data analysis in practice is an art as well as a science. Although statistical theory has foundations based on precise assumptions and conditions, in practice the real world is not so simple. This dichotomy makes it difficult to give gener-

al rules, always to be obeyed, for analyzing data. To that end, we have embedded numerous in-text references and separate ***In Practice*** boxes that alert students to the way statisticians actually analyze data in practice. These comments are based on our extensive consulting experience and by observing what well-trained statisticians do in practice.

### Activities and Simulations for Hands-On Learning

Activities are an important pedagogical tool for promoting understanding by helping the student to learn statistical concepts through hands-on exploration. Although we believe the activities help to facilitate learning, we have made them optional (and placed most of them at the ends of the chapters) so that instructors may use them at their discretion. Each of Chapters 1–11 includes at least one activity. A distinction is made between those activities that are to be done in class and those activities that are to be done by students outside of class. The activity often involves simulation, commonly using an applet available on the text CD and within MyMathLab. Additional activities are available on the text CD and MyMathLab. We encourage the use of these activities and simulations to promote student learning.

### Clearly Presented Definition and Summary Boxes

We highlight key definitions, guidelines, procedures, “In Practice” remarks, and other summaries in boxes throughout the text.

### Immediate Help for the Student

Material presented in the margins or boxed off from the main text offers at-a-glance information to aid in comprehending material in the main text. During our class tests, students have given high ratings to these features in making the book user-friendly. The margin features include the following

- **In Words:** This box explains the definitions and symbolic notation found in the body of the text itself (which, for technical accuracy, must be more formal).
- **Recall:** As the student progresses through the book, concepts are presented that call back to information learned in previous chapters. The Recall box directs the reader back to a previous presentation in the text to review and reinforce concepts and methods already covered.
- **Activity/Applet:** The reader is directed to an activity or applet on the text CD or within MyMathLab. Applets have great value, because they vividly show students certain concepts visually. For example, creating a sampling distribution is accomplished more readily with applets than with a static text figure. As noted, the activities and applets are presented as optional explorations in the text.

### Enhanced Visuals and a Graphical Approach to Learning Statistics

Increasingly, students are visual learners. Recognizing this fact, we have taken extra care to make the text figures informative. For instance, we’ve annotated many of the figures with labels that clearly identify the noteworthy aspects of the illustration. Further, most figure captions include a question (answered at the end of the chapter) designed to challenge the student to interpret and think about the information being communicated by the graphic. The graphics also feature a pedagogical use of color to help students recognize patterns and to distinguish between statistics and parameters. The use of color is explained in the inside front cover for easy access by students.



## Up-to-Date Use of Technology

The ready availability of technology helps instructors to provide instruction that is less calculation based and more concepts oriented. Output from *computer software and calculators* is displayed throughout the textbook. The discussion focuses on interpretation of the output, rather than on the keystrokes needed to create the output. We use the generic-looking, popular, and reliable MINITAB statistical software for most of our output examples but also include selected screen captures from the TI-83+/84 graphing calculator, SPSS, and Microsoft EXCEL, as appropriate. Technology specific manuals are also available with this text (MINITAB, SPSS, Excel, and TI-83+/84). See the supplements listing within the front matter for more information.

## An Invitation Rather Than a Conclusion

We hope that students using this textbook will gain a lasting appreciation for the vital role the art and science of statistics plays in analyzing data and helping us make decisions in our lives. As evidenced by the approach we've outlined above, our major goals for this textbook are that students should learn how:

- To produce data that can provide answers to properly posed questions
- To appreciate how probability helps us understand randomness in our lives, as well as to grasp the crucial concept of a sampling distribution and how it relates to inference methods
- To choose appropriate descriptive and inferential methods for examining and analyzing data and drawing conclusions
- To communicate the conclusions of statistical analyses clearly and effectively
- To understand the limitations of most research, whether because it was based on an observational study rather than a randomized experiment or survey, or because a certain lurking variable was not measured that could have explained the observed associations

We are excited about sharing through this text the insights that we have learned from our experience as teachers and from our students. Many students still enter statistics classes on the first day with dread because of its reputation as a dry, sometimes difficult, course. It is our goal to inspire a classroom environment that is filled with creativity, openness, realistic applications, and learning that students find inviting and rewarding. Many of our students have elected to take our second courses in statistics and even major or minor in statistics. We hope that this textbook will help the instructor and the students to experience a rewarding introductory course in statistics.

Alan Agresti,  
*University of Florida, Gainesville*

Christine Franklin,  
*University of Georgia, Athens*

## Acknowledgments

We are indebted to the many reviewers, class testers, and students who gave us invaluable feedback and advice on how to improve the quality of the book. Many of our reviewers worked with us through several drafts of the manuscript and provided us with valuable insight into how to improve the manuscript at both a micro and macro level. They include the following individuals as well as others who prefer to remain anonymous:

**ARIZONA** Russel Carlson, The University of Arizona; Peter Flanagan-Hyde, Phoenix Country Day School • **CALIFORNIA** Robert Gould, University of

California Los Angeles; Dawn Holmes, University of California Santa Barbara; Rebecca Head, Bakersfield College; Susan Herring, Sonoma State University; Colleen Kelly, San Diego State University; Elaine McDonald, Sonoma State University; Helen Noble, San Diego State University • **CONNECTICUT** Anne Doyle, University of Connecticut; Dan Miller, Central Connecticut State University; Kathleen McLaughlin, University of Connecticut; Stephen Sawin, Fairfield University • **DISTRICT OF COLUMBIA** Hans Engler, Georgetown University, Washington D.C. • **FLORIDA** James Lang, Valencia Community College; Megan Meece, University of Florida; Maria Ripol, University of Florida • **GEORGIA** Carrie Chmielarski, Student The University of Georgia; Charles LeMarsh, Lakeside High School; Broderick Oluyede, Georgia Southern University; Chandler Pike, The University of Georgia; Kim Robinson, Clayton State University; Jill Smith, University of Georgia; John Seppala, Valdosta State University; Joseph Walker, Georgia State University • **IOWA** John Cryer, The University of Iowa; William Duckworth, Iowa State University • **ILLINOIS** Jinadasa Gamage, Illinois State University; Richard Maher, Loyola University Chicago; Cathy Poliak, Northern Illinois University • **MASSACHUSETTS** Katherine Halvorsen, Smith College; Xiao-Li Meng, Harvard University; Daniel Weiner, Boston University • **MICHIGAN** Kirk Anderson, Grand Valley State University; Phyllis Curtiss, Grand Valley State University; Alla Sikorskii, Michigan State University • **MINNESOTA** Bob Dobrow, Carleton College • **MISSOURI** Larry Ries, University of Missouri–Columbia • **MONTANA** Jeff Banfield, Montana State University • **NORTH CAROLINA** Pamela Arroyo, North Carolina State University; E. Jacquelin Dietz, North Carolina State University; Roger Woodard, North Carolina State University • **NEBRASKA** Linda Young, University of Nebraska • **NEW JERSEY** Harold Sackrowitz, Rutgers, The State University of New Jersey • **NEW YORK** Brooke Fridley, Mohawk Valley Community College; Martin Lindquist, Columbia University; David Mathiason, Rochester Institute of Technology; Tian Zheng, Columbia University • **OHIO** John Holcomb, Cleveland State University; Jackie Miller, The Ohio State University; Stephan Pelikan, University of Cincinnati; Deborah Rumsey, The Ohio State University; Kevin Robinson, University of Akron • **OREGON** Henry Mesa, Portland Community College, Rock Creek; Qi-Man Shao, University of Oregon; Daming Xu, University of Oregon • **PENNSYLVANIA** Steven Gendler, Clarion University; Paul Lupinacci, Villanova University; Linda Myers, Harrisburg Area Community College • **SOUTH CAROLINA** Martin Jones, College of Charleston; Murray Siegel, The South Carolina Governor's School for Science and Mathematics; Webster West, University of South Carolina • **TENNESSEE** Robert Price, East Tennessee State University; Ginger Rowell, Middle Tennessee State University; Edith Seier, East Tennessee State University • **TEXAS** Larry Ammann, The University of Texas at Dallas; Tom Bratcher, Baylor University; Jianguo Liu, University of North Texas; Mary Parker, Austin Community College; Robert Paige, Texas Tech University; Diane Resnick, University of Houston–Downtown • **UTAH** Carolyn Cuff, Westminster College; Lajos Horvath, The University of Utah • **VIRGINIA** David Bauer, Virginia Commonwealth University; Ching-Yuan Chiang, James Madison University; Steven Garren, James Madison University; Debra Hydorn, Mary Washington College; D'Arcy Mays, Virginia Commonwealth University; Stephanie Pickle, Virginia Polytechnic Institute and State University • **WASHINGTON** J.R. Alldredge, Washington State University • **WYOMING** Burke Grandjean, University of Wyoming • **CANADA** David Loewen, University of Manitoba, Manitoba

Special thanks is due to the participants in our Chicago reviewer conference who spent two full days helping us think through the challenges associated with writing a first edition. They are Larry Amman, Phyllis Curtiss, Steve Garren, Martin Jones, James Lang, David Loewen, Megan Meece, Steve Pelikan, and Roger Woodard.



## Class Testers:

We must also single out those individuals who braved an untested manuscript and asked their students to help work out the kinks in the manuscript through our many class tests:

### ARIZONA

Peter Flanagan-Hyde, Phoenix Country Day School

### CALIFORNIA

Amy McElroy, Corey Manchester, San Diego State University

Calvin Schmall, Solano Community College

### COLORADO

David Most, Colorado State University

### CONNECTICUT

Paul Bugl, University of Hartford

Dan Miller, Central Connecticut State University

Steve Sawin, Fairfield University

John Vangor, Fairfield University

### FLORIDA

Nazanin Azarnia, Santa Fe Community College

### GEORGIA

Ouida Dillon, Oconee County High School

Katherine Hawks, Meadowcreek High School

Charles LaMarsh, Lakeside High School

Steve Messig, Oconee County High School

Broderick Oluyede, Georgia Southern University

Chandler Pike, University of Georgia

John Seppala, Valdosta State University

### MICHIGAN

Roy Erickson, Michigan State University

Mark Stevenson, Oakland Community College

### NEVADA

Alison Davis, University of Nevada, Reno

### NEW YORK

Steve Stehman, SUNY ESF

### PENNSYLVANIA

Bonnie A. Green, East Stroudsburg University

Deborah Lurie, Saint Joseph's University

### SOUTH CAROLINA

Martin Jones, Beverly Diamond, College of Charleston

Murray Siegel, The South Carolina Governor's School for Science and Mathematics

### TENNESSEE

Bonney Daves, Christian Academy of Knoxville

Ginger Holmes Rowell, Middle Tennessee State University

### VIRGINIA

Debra Hydorn, University of Mary Washington

### CANADA

David Loewen, University of Manitoba

**The following students also helped by way of a class test or focus group:**

### CONNECTICUT:

**Central Connecticut State University**—Chris Castaldo, Niki Chieka, Melissa Rocco, Angelita K. Tolly, Bryan Tracey, Jaime Valentine