

Introduction to the Practice of
STATISTICS NINTH
EDITION

Moore / McCabe / Craig

Introduction to the Practice of Statistics

NINTH EDITION

David S. Moore

George P. McCabe

Bruce A. Craig

Purdue University



w. h. freeman

Macmillan Learning

New York

Vice President, STEM: Ben Roberts
Publisher: Terri Ward
Senior Acquisitions Editor: Karen Carson
Marketing Manager: Tom DeMarco
Marketing Assistant: Cate McCaffery
Development Editor: Jorge Amaral
Senior Media Editor: Catriona Kaplan
Assistant Media Editor: Emily Tenenbaum
Director of Digital Production: Keri deManigold
Senior Media Producer: Alison Lorber
Associate Editor: Victoria Garvey
Editorial Assistant: Katharine Munz
Photo Editor: Cecilia Varas
Photo Researcher: Candice Cheesman
Director of Design, Content Management: Diana Blume
Text and Cover Designer: Blake Logan
Project Editor: Edward Dionne, MPS North America LLC
Illustrations: MPS North America LLC
Production Manager: Susan Wein
Composition: MPS North America LLC
Printing and Binding: LSC Communications
Cover Illustration: Drawing Water: Spring 2011 detail
(Midwest) by David Wicks
"Look Back" Arrow: NewCorner/Shutterstock

Library of Congress Control Number: 2016946039

Student Edition Hardcover:
ISBN-13: 978-1-319-01338-7
ISBN-10: 1-319-01338-4

Student Edition Loose-leaf:
ISBN-13: 978-1-319-01362-2
ISBN-10: 1-319-01362-7

Instructor Complimentary Copy:
ISBN-13: 978-1-319-01428-5
ISBN-10: 1-319-01428-3

© 2017, 2014, 2012, 2009 by W. H. Freeman and Company
All rights reserved
Printed in the United States of America
First printing

W. H. Freeman and Company
One New York Plaza
Suite 4500
New York, NY 10004-1562
www.macmillanlearning.com

Brief Contents

To Teachers: About This Book
To Students: What Is Statistics?
About the Authors
Data Table Index
Beyond the Basics Index

PART I Looking at Data

CHAPTER 1 Looking at Data—Distributions

CHAPTER 2 Looking at Data—Relationships

CHAPTER 3 Producing Data

PART II Probability and Inference

CHAPTER 4 Probability: The Study of Randomness

CHAPTER 5 Sampling Distributions

CHAPTER 6 Introduction to Inference

CHAPTER 7 Inference for Means

CHAPTER 8 Inference for Proportions

PART III Topics in Inference

CHAPTER 9 Inference for Categorical Data

CHAPTER 10 Inference for Regression

CHAPTER 11 Multiple Regression

CHAPTER 12 One-Way Analysis of Variance

CHAPTER 13 Two-Way Analysis of Variance

Tables

Answers to Odd-Numbered Exercises

Notes and Data Sources

Index

Contents

To Teachers: About This Book
To Students: What Is Statistics?
About the Authors
Data Table Index
Beyond the Basics Index

PART I Looking at Data

CHAPTER 1 Looking at Data—Distributions

Introduction

1.1 Data

Key characteristics of a data set

Section 1.1 Summary

Section 1.1 Exercises

1.2 Displaying Distributions with Graphs

Categorical variables: Bar graphs and pie charts

Quantitative variables: Stemplots and histograms

Histograms

Data analysis in action: Don't hang up on me

Examining distributions

Dealing with outliers

Time plots

Section 1.2 Summary

Section 1.2 Exercises

1.3 Describing Distributions with Numbers

Measuring center: The mean

Measuring center: The median

Mean versus median

Measuring spread: The quartiles

The five-number summary and boxplots

The $1.5 \times IQR$ rule for suspected outliers

Measuring spread: The standard deviation

Properties of the standard deviation

Choosing measures of center and spread

Changing the unit of measurement

Section 1.3 Summary

Section 1.3 Exercises

1.4 Density Curves and Normal Distributions

Density curves

Measuring center and spread for density curves

Normal distributions

The 68–95–99.7 rule

Standardizing observations

Normal distribution calculations

Using the standard Normal table

Inverse Normal calculations

Normal quantile plots

Beyond the Basics: Density estimation

Section 1.4 Summary

Section 1.4 Exercises

Chapter 1 Exercises

CHAPTER 2 Looking at Data—Relationships

Introduction

2.1 Relationships

Examining relationships

Section 2.1 Summary

Section 2.1 Exercises

2.2 Scatterplots

Interpreting scatterplots

The log transformation

Adding categorical variables to scatterplots

Scatterplot smoothers

Categorical explanatory variables

Section 2.2 Summary

Section 2.2 Exercises

2.3 Correlation

The correlation r

Properties of correlation

Section 2.3 Summary

Section 2.3 Exercises

2.4 Least-Squares Regression

Fitting a line to data

Prediction

- Least-squares regression
- Interpreting the regression line
- Facts about least-squares regression
- Correlation and regression
- Another view of r^2

Section 2.4 Summary

Section 2.4 Exercises

2.5 Cautions about Correlation and Regression

- Residuals
- Outliers and influential observations
- Beware of the lurking variable
- Beware of correlations based on averaged data
- Beware of restricted ranges

Beyond the Basics: Data mining

Section 2.5 Summary

Section 2.5 Exercises

2.6 Data Analysis for Two-Way Tables

- The two-way table
- Joint distribution
- Marginal distributions
- Describing relations in two-way tables
- Conditional distributions
- Simpson's paradox

Section 2.6 Summary

Section 2.6 Exercises

2.7 The Question of Causation

- Explaining association
- Establishing causation

Section 2.7 Summary
Section 2.7 Exercises
Chapter 2 Exercises

CHAPTER 3 Producing Data

Introduction

3.1 Sources of Data

- Anecdotal data
- Available data
- Sample surveys and experiments

Section 3.1 Summary

Section 3.1 Exercises

3.2 Design of Experiments

- Comparative experiments
- Randomization
- Randomized comparative experiments
- How to randomize
- Randomization using software
- Randomization using random digits
- Cautions about experimentation
- Matched pairs designs
- Block designs

Section 3.2 Summary

Section 3.2 Exercises

3.3 Sampling Design

- Simple random samples
- How to select a simple random sample

- Stratified random samples
- Multistage random samples
- Cautions about sample surveys

[Beyond the Basics: Capture-recapture sampling](#)

[Section 3.3 Summary](#)

[Section 3.3 Exercises](#)

3.4 Ethics

- Institutional review boards
- Informed consent
- Confidentiality
- Clinical trials
- Behavioral and social science experiments

[Section 3.4 Summary](#)

[Section 3.4 Exercises](#)

[Chapter 3 Exercises](#)

PART II Probability and Inference

CHAPTER 4 Probability: The Study of Randomness

Introduction

4.1 Randomness

- The language of probability
- Thinking about randomness
- The uses of probability

[Section 4.1 Summary](#)

[Section 4.1 Exercises](#)

4.2 Probability Models

Sample spaces

Probability rules

Assigning probabilities: Finite number of outcomes

Assigning probabilities: Equally likely outcomes

Independence and the multiplication rule

Applying the probability rules

Section 4.2 Summary

Section 4.2 Exercises

4.3 Random Variables

Discrete random variables

Continuous random variables

Normal distributions as probability distributions

Section 4.3 Summary

Section 4.3 Exercises

4.4 Means and Variances of Random Variables

The mean of a random variable

Statistical estimation and the law of large numbers

Thinking about the law of large numbers

Beyond the Basics: More laws of large numbers

Rules for means

The variance of a random variable

Rules for variances and standard deviations

Section 4.4 Summary

Section 4.4 Exercises

4.5 General Probability Rules

General addition rules

- Conditional probability
- General multiplication rules
- Tree diagrams
- Bayes's rule
- Independence again

Section 4.5 Summary

Section 4.5 Exercises

Chapter 4 Exercises

CHAPTER 5 Sampling Distributions

Introduction

5.1 Toward Statistical Inference

- Sampling variability
- Sampling distributions
- Bias and variability
- Sampling from large populations
- Why randomize?

Section 5.1 Summary

Section 5.1 Exercises

5.2 The Sampling Distribution of a Sample Mean

- The mean and standard deviation of $\bar{x}$
- The central limit theorem
- A few more facts

Beyond the Basics: Weibull distributions

Section 5.2 Summary

Section 5.2 Exercises

5.3 Sampling Distributions for Counts and Proportions

- The binomial distributions for sample counts
- Binomial distributions in statistical sampling
- Finding binomial probabilities
- Binomial mean and standard deviation
- Sample proportions
- Normal approximation for counts and proportions
- The continuity correction
- Binomial formula
- The Poisson distributions

Section 5.3 Summary

Section 5.3 Exercises

Chapter 5 Exercises

CHAPTER 6 Introduction to Inference

Introduction

Overview of inference

6.1 Estimating with Confidence

- Statistical confidence
- Confidence intervals
- Confidence interval for a population mean
- How confidence intervals behave
- Choosing the sample size
- Some cautions

Section 6.1 Summary

Section 6.1 Exercises

6.2 Tests of Significance

The reasoning of significance tests

Stating hypotheses

Test statistics

P -values

Statistical significance

Tests for a population mean

Two-sided significance tests and confidence intervals

The P -value versus a statement of significance

Section 6.2 Summary

Section 6.2 Exercises

6.3 Use and Abuse of Tests

Choosing a level of significance

What statistical significance does not mean

Don't ignore lack of significance

Statistical inference is not valid for all sets of data

Beware of searching for significance

Section 6.3 Summary

Section 6.3 Exercises

6.4 Power and Inference as a Decision

Power

Increasing the power

Inference as decision

Two types of error

Error probabilities

The common practice of testing hypotheses

Section 6.4 Summary
Section 6.4 Exercises
Chapter 6 Exercises

CHAPTER 7 Inference for Means

Introduction

7.1 Inference for the Mean of a Population

- The t distributions
- The one-sample t confidence interval
- The one-sample t test
- Matched pairs t procedures
- Robustness of the t procedures

Beyond the Basics: The bootstrap

Section 7.1 Summary
Section 7.1 Exercises

7.2 Comparing Two Means

- The two-sample z statistic
- The two-sample t procedures
- The two-sample t confidence interval
- The two-sample t significance test
- Robustness of the two-sample procedures
- Inference for small samples
- Software approximation for the degrees of freedom
- The pooled two-sample t procedures

Section 7.2 Summary
Section 7.2 Exercises

7.3 Additional Topics on Inference

Choosing the sample size

Inference for non-Normal populations

Section 7.3 Summary

Section 7.3 Exercises

Chapter 7 Exercises

CHAPTER 8 Inference for Proportions

Introduction

8.1 Inference for a Single Proportion

Large-sample confidence interval for a single proportion

Beyond the Basics: The plus four confidence interval for a single proportion

Significance test for a single proportion

Choosing a sample size for a confidence interval

Choosing a sample size for a significance test

Section 8.1 Summary

Section 8.1 Exercises

8.2 Comparing Two Proportions

Large-sample confidence interval for a difference in proportions

Beyond the Basics: The plus four confidence interval for a difference in proportions

Significance test for a difference in proportions

Choosing a sample size for two sample proportions

Beyond the Basics: Relative risk

Section 8.2 Summary

Section 8.2 Exercises

Chapter 8 Exercises

PART III Topics in Inference

CHAPTER 9 Inference for Categorical Data

Introduction

9.1 Inference for Two-Way Tables

- The hypothesis: No association

- Expected cell counts

- The chi-square test

- Computations

- Computing conditional distributions

- The chi-square test and the z test

Beyond the Basics: Meta-analysis

Section 9.1 Summary

Section 9.1 Exercises

9.2 Goodness of Fit

Section 9.2 Summary

Section 9.2 Exercises

Chapter 9 Exercises

CHAPTER 10 Inference for Regression

Introduction

10.1 Simple Linear Regression

Statistical model for linear regression
Preliminary data analysis and inference considerations
Estimating the regression parameters
Checking model assumptions
Confidence intervals and significance tests
Confidence intervals for mean response
Prediction intervals
Transforming variables

Beyond the Basics: Nonlinear regression

Section 10.1 Summary

Section 10.1 Exercises

10.2 More Detail about Simple Linear Regression

Analysis of variance for regression
The ANOVA F test
Calculations for regression inference
Inference for correlation

Section 10.2 Summary

Section 10.2 Exercises

Chapter 10 Exercises

CHAPTER 11 Multiple Regression

Introduction

11.1 Inference for Multiple Regression

Population multiple regression equation
Data for multiple regression
Multiple linear regression model

Estimation of the multiple regression parameters

Confidence intervals and significance tests for regression coefficients

ANOVA table for multiple regression

Squared multiple correlation R^2

Section 11.1 Summary

Section 11.1 Exercises

11.2 A Case Study

Preliminary analysis

Relationships between pairs of variables

Regression on high school grades

Interpretation of results

Examining the residuals

Refining the model

Regression on SAT scores

Regression using all variables

Test for a collection of regression coefficients

Beyond the Basics: Multiple logistic regression

Section 11.2 Summary

Section 11.2 Exercises

Chapter 11 Exercises

CHAPTER 12 One-Way Analysis of Variance

Introduction

12.1 Inference for One-Way Analysis of Variance

Data for one-way ANOVA

Comparing means

The two-sample t statistic
An overview of ANOVA
The ANOVA model
Estimates of population parameters
Testing hypotheses in one-way ANOVA
The ANOVA table
The F test
Software

Beyond the Basics: Testing the equality of spread

Section 12.1 Summary

Section 12.1 Exercises

12.2 Comparing the Means

Contrasts

Multiple comparisons

Power

Section 12.2 Summary

Section 12.2 Exercises

Chapter 12 Exercises

CHAPTER 13 Two-Way Analysis of Variance

Introduction

13.1 The Two-Way ANOVA Model

Advantages of two-way ANOVA

The two-way ANOVA model

Main effects and interactions

13.2 Inference for Two-Way ANOVA

The ANOVA table for two-way ANOVA

Chapter 13 Summary

Chapter 13 Exercises

Tables

Answers to Odd-Numbered Exercises

Notes and Data Sources

Index

To Teachers: About This Book

Statistics is the science of data. *Introduction to the Practice of Statistics (IPS)* is an introductory text based on this principle. We present methods of basic statistics in a way that emphasizes working with data and mastering statistical reasoning. *IPS* is elementary in mathematical level but conceptually rich in statistical ideas. After completing a course based on our text, we would like students to be able to think objectively about conclusions drawn from data and use statistical methods in their own work.

In *IPS*, we combine attention to basic statistical concepts with a comprehensive presentation of the elementary statistical methods that students will find useful in their work. *IPS* has been successful for several reasons:

1. *IPS* examines the nature of modern statistical practice at a level suitable for beginners. We focus on the production and analysis of data as well as the traditional topics of probability and inference.
2. *IPS* has a logical overall progression, so data production and data analysis are a major focus, while inference is treated as a tool that helps us draw conclusions from data in an appropriate way.
3. *IPS* presents data analysis as more than a collection of techniques for exploring data. We emphasize systematic ways of thinking about data. Simple principles guide the analysis: always plot your data; look for overall patterns and deviations from them; when looking at the overall pattern of a distribution for one variable, consider shape, center, and spread; for relations between two variables, consider form, direction, and

strength; always ask whether a relationship between variables is influenced by other variables lurking in the background. We warn students about pitfalls in clear cautionary discussions.

4. *IPS* uses real examples to drive the exposition. Students learn the technique of least-squares regression and how to interpret the regression slope. But they also learn the conceptual ties between regression and correlation and the importance of looking for influential observations.
5. *IPS* is aware of current developments both in statistical science and in teaching statistics. Brief, optional Beyond the Basics sections give quick overviews of topics such as density estimation, scatterplot smoothers, data mining, nonlinear regression, and meta-analysis. Chapter 16 gives an elementary introduction to the bootstrap and other computer-intensive statistical methods.

The title of the book expresses our intent to introduce readers to statistics as it is used in practice. Statistics in practice is concerned with drawing conclusions from data. We focus on problem solving rather than on methods that may be useful in specific settings.

GAISE The College Report of the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project (www.amstat.org/education/gaise/) was funded by the American Statistical Association to make recommendations for how introductory statistics courses should be taught. This report and its update contain many interesting teaching suggestions, and we strongly recommend that you read it. The philosophy and approach of *IPS* closely reflect the GAISE recommendations. Let's examine each of the latest recommendations in the context of *IPS*.

1. *Teach statistical thinking.* Through our experiences as applied statisticians, we are very familiar with the components that are needed for the appropriate use of statistical methods. We focus on formulating questions, collecting and finding data, evaluating the quality of data, exploring the relationships among variables, performing statistical analyses, and drawing conclusions. In examples and exercises throughout the text, we emphasize

putting the analysis in the proper context and translating numerical and graphical summaries into conclusions.

2. *Focus on conceptual understanding.* With the software available today, it is very easy for almost anyone to apply a wide variety of statistical procedures, both simple and complex, to a set of data. Without a firm grasp of the concepts, such applications are frequently meaningless. By using the methods that we present on real sets of data, we believe that students will gain an excellent understanding of these concepts. Our emphasis is on the input (questions of interest, collecting or finding data, examining data) and the output (conclusions) for a statistical analysis. Formulas are given only where they will provide some insight into concepts.
3. *Integrate real data with a context and a purpose.* Many of the examples and exercises in *IPS* include data that we have obtained from collaborators or consulting clients. Other data sets have come from research related to these activities. We have also used the Internet as a data source, particularly for data related to social media and other topics of interest to undergraduates. Our emphasis on real data, rather than artificial data chosen to illustrate a calculation, serves to motivate students and help them see the usefulness of statistics in everyday life. We also frequently encounter interesting statistical issues that we explore. These include outliers and nonlinear relationships. All data sets are available from the text website.
4. *Foster active learning in the classroom.* As we mentioned earlier, we believe that statistics is exciting as something to do rather than something to talk about. Throughout the text, we provide exercises in Use Your Knowledge sections that ask the students to perform some relatively simple tasks that reinforce the material just presented. Other exercises are particularly suited to being worked on and discussed within a classroom setting.
5. *Use technology for developing concepts and analyzing data.* Technology has altered statistical practice in a fundamental way. In the past, some of the calculations that we performed were

particularly difficult and tedious. In other words, they were not fun. Today, freed from the burden of computation by software, we can concentrate our efforts on the big picture: what questions are we trying to address with a study and what can we conclude from our analysis?

6. *Use assessments to improve and evaluate student learning.* Our goal for students who complete a course based on *IPS* is that they are able to design and carry out a statistical study for a project in their capstone course or other setting. Our exercises are oriented toward this goal. Many ask about the design of a statistical study and the collection of data. Others ask for a paragraph summarizing the results of an analysis. This recommendation includes the use of projects, oral presentations, article critiques, and written reports. We believe that students using this text will be well prepared to undertake these kinds of activities. Furthermore, we view these activities not only as assessments but also as valuable tools for learning statistics.

Teaching Recommendations We have used *IPS* in courses taught to a variety of student audiences. For general undergraduates from mixed disciplines, we recommend covering [Chapters 1](#) through [8](#) and [Chapters 9](#), [10](#), or [12](#). For a quantitatively strong audience—sophomores planning to major in actuarial science or statistics—we recommend moving more quickly. Add [Chapters 10](#) and [11](#) to the core material in [Chapters 1](#) through [8](#). In general, we recommend deemphasizing the material on probability because these students will take a probability course later in their program. For beginning graduate students in such fields as education, family studies, and retailing, we recommend that the students read the entire text ([Chapters 11](#) and [13](#) lightly), again with reduced emphasis on [Chapter 4](#) and some parts of [Chapter 5](#). In all cases, beginning with data analysis and data production ([Part I](#)) helps students overcome their fear of statistics and builds a sound base for studying inference. We believe that *IPS* can easily be adapted to a wide variety of audiences.

The Ninth Edition: What's New?

- [Chapter 1](#) now begins with a short section giving an overview of data.
- “Toward Statistical Inference” (previously [Section 3.3](#)), which introduces the concepts of statistical inference and sampling distributions, has been moved to [Section 5.1](#) to better assist with the transition from a single data set to sampling distributions.
- Coverage of mosaic plots as a visual tool for relationships between two categorical variables has been added to [Chapters 2](#) and [9](#).
- [Chapter 3](#) now begins with a short section giving a basic overview of data sources.
- Coverage of equivalence testing has been added to [Chapter 7](#).
- There is a greater emphasis on sample size determination using software in [Chapters 7](#) and [8](#).
- Resampling and bootstrapping are now introduced in [Chapter 7](#) rather than [Chapter 6](#).
- “Inference for Categorical Data” is the new title for [Chapter 9](#), which includes goodness of fit as well as inference for two-way tables.
- There are more JMP screenshots and updated screenshots of Minitab, Excel, and SPSS outputs.
- **Design** A new design incorporates colorful, revised figures throughout to aid the students’ understanding of text material. Photographs related to chapter examples and exercises make connections to real-life applications and provide a visual context for topics. More figures with software output have been included.
- **Exercises and Examples** More than 30% of the exercises are new or revised, and there are more than 1700 exercises total. Exercise sets have been added at the end of sections in [Chapters 9](#) through [12](#). To maintain the attractiveness of the examples to students, we have replaced or updated a large

number of them. More than 30% of the 430 examples are new or revised. A list of exercises and examples categorized by application area is provided on the inside of the front cover.

In addition to the new ninth edition enhancements, *IPS* has retained the successful pedagogical features from previous editions:



- **Look Back** At key points in the text, Look Back margin notes direct the reader to the first explanation of a topic, providing page numbers for easy reference.



- **Caution** Warnings in the text, signaled by a caution icon, help students avoid common errors and misconceptions.



- **Challenge Exercises** More challenging exercises are signaled with an icon. Challenge exercises are varied: some are mathematical, some require open-ended investigation, and others require deeper thought about the basic concepts.



- **Applets** Applet icons are used throughout the text to signal where related interactive statistical applets can be found on the *IPS* website and in LaunchPad.

- **Use Your Knowledge Exercises** We have found these exercises to be a very useful learning tool. They appear throughout each section and are listed, with page numbers, before the section-ending exercises.
- **Technology output screenshots** Most statistical analyses rely heavily on statistical software. In this book, we discuss the use of Excel 2013, JMP 12, Minitab 17, SPSS 23, CrunchIt, R, and a TI-83/-84 calculator for conducting statistical analysis. As specialized statistical packages, JMP, Minitab, and SPSS are the most popular software choices both in industry and in colleges and schools of business. R is an extremely powerful statistical environment that is free to anyone; it relies heavily on members of the academic and general statistical communities for support. As an all-purpose spreadsheet program, Excel provides a limited set of statistical analysis options in comparison. However, given its pervasiveness and wide acceptance in industry and the computer world at large, we believe it is important to give Excel proper attention. It should be noted that for users who want more statistical capabilities but want to work in an Excel environment, there are a number of commercially available add-on packages (if you have JMP, for instance, it can be invoked from within Excel). Finally, instructions are provided for the TI-83/-84 calculators.

Even though basic guidance is provided in the book, it should be emphasized that *IPS* is not bound to any of these programs. Computer output from statistical packages is very similar, so you can feel quite comfortable using any one these packages.

Acknowledgments

We are pleased that the first eight editions of *Introduction to the Practice of Statistics* have helped to move the teaching of introductory statistics in a direction supported by most statisticians. We are grateful to the many colleagues and students who have

provided helpful comments, and we hope that they will find this new edition another step forward. In particular, we would like to thank the following colleagues who offered specific comments on the new edition:

Ali Arab, *Georgetown University*

Tessema Astatkie, *Dalhousie University*

Fouzia Baki, *McMaster University*

Lynda Ballou, *New Mexico Institute of Mining and Technology*

Sanjib Basu, *Northern Illinois University*

David Bosworth, *Hutchinson Community College*

Max Buot, *Xavier University*

Nadjib Bouzar, *University of Indianapolis*

Matt Carlton, *California Polytechnic State University–San Luis Obispo*

Gustavo Cepparo, *Austin Community College*

Pinyuen Chen, *Syracuse University*

Dennis L. Clason, *University of Cincinnati–Blue Ash College*

Tadd Colver, *Purdue University*

Chris Edwards, *University of Wisconsin–Oshkosh*

Irina Gaynanova, *Texas A&M University*

Brian T. Gill, *Seattle Pacific University*

Mary Gray, *American University*

Gary E. Haefner, *University of Cincinnati*

Susan Herring, *Sonoma State University*

Lifang Hsu, *Le Moyne College*

Tiffany Kolba, *Valparaiso University*

Lia Liu, *University of Illinois at Chicago*

Xuewen Lu, *University of Calgary*

Antoinette Marquard, *Cleveland State University*

Frederick G. Schmitt, *College of Marin*

James D. Stamey, *Baylor University*

Engin Sungur, *University of Minnesota–Morris*

Anatoliy Swishchuk, *University of Calgary*

Richard Tardanico, *Florida International University*

Melanee Thomas, *University of Calgary*

Terri Torres, *Oregon Institute of Technology*

Mahbobeh Vezvaei, *Kent State University*
Yishi Wang, *University of North Carolina–Wilmington*
John Ward, *Jefferson Community and Technical College*
Debra Wiens, *Rocky Mountain College*
Victor Williams, *Paine College*
Christopher Wilson, *Butler University*
Anne Yust, *Birmingham-Southern College*
Biao Zhang, *The University of Toledo*
Michael L. Zwilling, *University of Mount Union*

The professionals at Macmillan, in particular, Terri Ward, Karen Carson, Jorge Amaral, Emily Tenenbaum, Ed Dionne, Blake Logan, and Susan Wein, have contributed greatly to the success of *IPS*. In addition, we would like to thank Tadd Colver at Purdue University for his valuable contributions to the ninth edition, including authoring the back-of-book answers, solutions, and Instructor's Guide. We'd also like to thank Monica Jackson at American University for accuracy reviewing the back-of-book answers and solutions and for authoring the test bank. Thanks also to Michael Zwilling at University of Mount Union for accuracy reviewing the test bank, Christopher Edwards at University of Wisconsin Oshkosh for authoring the lecture slides, and James Stamey at Baylor University for authoring the Clicker slides.

Most of all, we are grateful to the many friends and collaborators whose data and research questions have enabled us to gain a deeper understanding of the science of data. Finally, we would like to acknowledge the contributions of John W. Tukey, whose contributions to data analysis have had such a great influence on us as well as a whole generation of applied statisticians.

Media and Supplements



LaunchPad, our online course space, combines an interactive e-Book with high-quality multimedia content and ready-made assessment options, including LearningCurve adaptive quizzing. Content is easy to assign or adapt with your own material, such as readings, videos, quizzes, discussion groups, and more. LaunchPad also provides access to a Gradebook that offers a window into your students' performance—either individually or as a whole. Use LaunchPad on its own or integrate it with your school's learning management system so your class is always on the same page. To learn more about LaunchPad for ***Introduction to the Practice of Statistics, Ninth Edition***, or to request access, go to launchpadworks.com.

Assets integrated into LaunchPad include:

Interactive e-Book. Every LaunchPad e-Book comes with powerful study tools for students, video and multimedia content, and easy customization for instructors. Students can search, highlight, and bookmark, making it easier to study and access key content. And teachers can ensure that their classes get just the book they want to deliver: customize and rearrange chapters; add and share notes and discussions; and link to quizzes, activities, and other resources.



LearningCurve provides students and instructors with powerful adaptive quizzing, a game-like format, direct links to the e-Book, and instant feedback. The quizzing system features questions tailored specifically to the text and adapts to students' responses, providing material at different difficulty levels and topics based on student performance.



JMP Student Edition (developed by SAS) is easy to learn and contains all the capabilities required for introductory statistics. JMP is the leading commercial data analysis software of choice for scientists, engineers, and analysts at companies throughout the world (for Windows and Mac). Register inside LaunchPad at no additional cost.



CrunchIt![®] is a Web-based statistical program that allows users to perform all the statistical operations and graphing needed for an introductory statistics course and more. It saves users time by automatically loading data from *IPS*, 9e, and it provides the flexibility to edit and import additional data.

StatBoards Videos are brief whiteboard videos that illustrate difficult topics through additional examples, written and explained by a select group of statistics educators.

Stepped Tutorials are centered on algorithmically generated quizzing with step-by-step feedback to help students work their way toward the correct solution. These exercise tutorials (two to three per chapter) are easily assignable and assessable.

Statistical Video Series consists of StatClips, StatClips Examples, and Statistically Speaking "Snapshots." View animated lecture videos, whiteboard lessons, and documentary-style footage that illustrate key statistical concepts and help students visualize statistics in real-world scenarios.

Video Technology Manuals, available for TI-83/84 calculators, Minitab, Excel, JMP, SPSS, R, Rcmdr, and CrunchIt![®], provide brief instructions for using specific statistical software.

StatTutor Tutorials offer multimedia tutorials that explore important concepts and procedures in a presentation that combines video, audio, and interactive features. The newly revised format includes built-in, assignable assessments and a bright new interface.

Statistical Applets give students hands-on opportunities to familiarize themselves with important statistical concepts and procedures in an interactive setting that allows them to manipulate variables and see the results graphically. Icons in the textbook indicate when an applet is available for the material being covered. Applets are assessable and assignable in LaunchPad.

Stats@Work Simulations put students in the role of the statistical consultant, helping them better understand statistics interactively within the context of real-life scenarios.

EESEE Case Studies (Electronic Encyclopedia of Statistical Examples and Exercises), developed by The Ohio State University Statistics Department, teach students to apply their statistical skills by exploring actual case studies using real data.



SolutionMaster offers an easy-to-use web-based version of the instructor's solutions, allowing instructors to generate a solution file for any set of homework exercises.

Data files are available in JMP, ASCII, Excel, TI, Minitab, SPSS (an IBM Company)*, R, and CSV formats.

Student Solutions Manual provides solutions to the odd-numbered exercises in the text and is available as a print supplement and electronically in LaunchPad.

Instructor's Guide with Full Solutions includes teaching suggestions, chapter comments, and detailed solutions to all exercises and is available electronically in LaunchPad.

Test Bank offers hundreds of multiple-choice questions and is available in LaunchPad.

Lecture Slides offer a customizable, detailed lecture presentation of statistical concepts covered in each chapter of *IPS*, 9e. **Image slides** contain all textbook figures and tables. Lecture slides and images slides are available in LaunchPad.



WebAssign offers algorithmic questions from *IPS*, 9e, in a powerful online instructional system. WebAssign lets you easily create assignments, grade homework, and give your students instant feedback. Along with flexible features, class and question-level analytics are available for instructors and students. WebAssign Premium also includes the following resources described above: e-Book, data files, LearningCurve, StatTutor Tutorials, Statistical Videos, Video Technology Manuals, solutions manuals, lecture and image slides, i-Clicker slides, test bank, and practice quizzes.

Additional Resources Available with *IPS*, 9e

Special Software Package A student version of JMP is available for packaging with the printed text. JMP is also available inside LaunchPad at no additional cost.



i-Clicker is a two-way radio-frequency classroom response solution developed by educators for educators. Each step of i-Clicker's development has been informed by teaching and learning.

* SPSS was acquired by IBM in October 2009

To Students: What Is Statistics?

Statistics is the science of collecting, organizing, and interpreting numerical facts, which we call *data*. We are bombarded by data in our everyday lives. The news mentions movie box-office sales, the latest poll of the president's popularity, and the average high temperature for today's date. Advertisements claim that data show the superiority of the advertiser's product. All sides in public debates about economics, education, and social policy argue from data. A knowledge of statistics helps separate sense from nonsense in this flood of data.

The study and collection of data are also important in the work of many professions, so training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisers, as well as policymakers in government and business, study these data in order to make informed decisions. Doctors must understand the origin and trustworthiness of the data that appear in medical journals. Politicians rely on data from polls of public opinion. Business decisions are based on market research data that reveal consumer tastes and preferences. Engineers gather data on the quality and reliability of manufactured products. Most areas of academic study make use of numbers and, therefore, also make use of the methods of statistics. This means it is extremely likely that your undergraduate research projects will involve, at some level, the use of statistics.

Learning from Data

The goal of statistics is to learn from data. To learn, we often perform calculations or make graphs based on a set of numbers. But to learn from data, we must do more than calculate and plot because data are not just numbers; they are numbers that have some context that helps us learn from them.

More than two-thirds of Americans are overweight or obese according to the Centers for Disease Control and Prevention (CDC) website (www.cdc.gov/nchs/nhanes.htm). What does it mean to be obese or to be overweight? To answer this question, we need to talk about body mass index (BMI). Your weight in kilograms divided by the square of your height in meters is your BMI. A man who is 6 feet tall (1.83 meters) and weighs 180 pounds (81.65 kilograms) will have a BMI of $81.65/(1.83)^2 = 24.4 \text{ kg/m}^2$. How do we interpret this number? According to the CDC, a person is classified as overweight if his or her BMI is between 25 and 29.9 kg/m^2 and as obese if his or her BMI is 30 kg/m^2 or more. Therefore, more than two-thirds of Americans have a BMI of 25 kg/m^2 or more. The man who weighs 180 pounds and is 6 feet tall is not overweight or obese, but if he gains 5 pounds, his BMI would increase to 25.1, and he would be classified as overweight.

When you do statistical problems, even straightforward textbook problems, don't just graph or calculate. Think about the context and state your conclusions in the specific setting of the problem. As you are learning how to do statistical calculations and graphs, remember that the goal of statistics is not calculation for its own sake but gaining understanding from numbers. The calculations and graphs can be automated by a calculator or software, but you must supply the understanding. This book presents only the most common specific procedures for statistical analysis. A thorough grasp of the principles of statistics will enable you to quickly learn more advanced methods as needed. On the other hand, a fancy computer analysis carried out without attention to basic principles will often produce elaborate nonsense. As you read, seek to understand the principles as well as the necessary details of methods and recipes.

The Rise of Statistics

Historically, the ideas and methods of statistics developed gradually as society grew interested in collecting and using data for a variety of applications. The earliest origins of statistics lie in the desire of rulers to count the number of inhabitants or measure the value of taxable land in their domains. As the physical sciences developed in the seventeenth and eighteenth centuries, the importance of careful measurements of weights, distances, and other physical quantities grew. Astronomers and surveyors striving for exactness had to deal with variation in their measurements. Many measurements should be better than a single measurement, even though they vary among themselves. How can we best combine many varying observations? Statistical methods that are still important were invented in order to analyze scientific measurements.

By the nineteenth century, the agricultural, life, and behavioral sciences also began to rely on data to answer fundamental questions. How are the heights of parents and children related? Does a new variety of wheat produce higher yields than the old, and under what conditions of rainfall and fertilizer? Can a person's mental ability and behavior be measured just as we measure height and reaction time? Effective methods for dealing with such questions developed slowly and with much debate.

As methods for producing and understanding data grew in number and sophistication, the new discipline of statistics took shape in the twentieth century. Ideas and techniques that originated in the collection of government data, in the study of astronomical or biological measurements, and in the attempt to understand heredity or intelligence came together to form a unified "science of data." That science of data—statistics—is the topic of this text.

The Organization of This Book

Part I of this book, called simply "Looking at Data," concerns data analysis and data production. The first two chapters deal with

statistical methods for organizing and describing data. These chapters progress from simpler to more complex data. [Chapter 1](#) examines data on a single variable; [Chapter 2](#) is devoted to relationships among two or more variables. You will learn both how to examine data produced by others and how to organize and summarize your own data. These summaries will first be graphical, then numerical, and then, when appropriate, in the form of a mathematical model that gives a compact description of the overall pattern of the data. [Chapter 3](#) outlines arrangements (called designs) for producing data that answer specific questions. The principles presented in this chapter will help you to design proper samples and experiments for your research projects and to evaluate other such investigations in your field of study.

[Part II](#), consisting of [Chapters 4](#) through [8](#), introduces statistical inference—formal methods for drawing conclusions from properly produced data. Statistical inference uses the language of probability to describe how reliable its conclusions are, so some basic facts about probability are needed to understand inference. Probability is the subject of [Chapters 4](#) and [5](#). [Chapter 6](#), perhaps the most important chapter in the text, introduces the reasoning of statistical inference. Effective inference is based on good procedures for producing data ([Chapter 3](#)), careful examination of the data ([Chapters 1](#) and [2](#)), and an understanding of the nature of statistical inference as discussed in [Chapter 6](#). [Chapters 7](#) and [8](#) describe some of the most common specific methods of inference, for drawing conclusions about means and proportions from one and two samples.

The five shorter chapters in [Part III](#) introduce somewhat more advanced methods of inference, dealing with relations in categorical data, regression and correlation, and analysis of variance. Four supplementary chapters, available from the text website, present additional statistical topics.

What Lies Ahead

Introduction to the Practice of Statistics is full of data from many different areas of life and study. Many exercises ask you to express briefly some understanding gained from the data. In practice, you would know much more about the background of the data you work with and about the questions you hope the data will answer. No textbook can be fully realistic. But it is important to form the habit of asking, “What do the data tell me?” rather than just concentrating on making graphs and doing calculations.

You should have some help in automating many of the graphs and calculations. You should certainly have a calculator with basic statistical functions. Look for keywords such as “two-variable statistics” or “regression” when you shop for a calculator. More advanced (and more expensive) calculators will do much more, including some statistical graphs. You may be asked to use software as well. There are many kinds of statistical software, from spreadsheets to large programs for advanced users of statistics. The kind of computing available to learners varies a great deal from place to place—but the big ideas of statistics don’t depend on any particular level of access to computing.

Because graphing and calculating are automated in statistical practice, the most important assets you can gain from the study of statistics are an understanding of the big ideas and the beginnings of good judgment in working with data. Ideas and judgment can’t (at least yet) be automated. They guide you in telling the computer what to do and in interpreting its output. This book tries to explain the most important ideas of statistics, not just teach methods. Some examples of big ideas that you will meet are “always plot your data,” “randomized comparative experiments,” and “statistical significance.”

You learn statistics by doing statistical problems. “Practice, practice, practice.” Be prepared to work problems. The basic principle of learning is persistence. Being organized and persistent is more helpful in reading this book than knowing lots of math. 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 gain will be worth the pain.

About the Authors

David S. Moore is Shanti S. Gupta Distinguished Professor of Statistics, Emeritus, at Purdue University and was 1998 president of the American Statistical Association. He received his AB from Princeton and his PhD from Cornell, both in mathematics. He has written many research papers in statistical theory and served on the editorial boards of several major journals.

Professor Moore is an elected fellow of the American Statistical Association and of the Institute of Mathematical Statistics and is an elected member of the International Statistical Institute. He has served as program director for statistics and probability at the National Science Foundation.

In recent years, Professor Moore has devoted his attention to the teaching of statistics. He was the content developer for the Annenberg/Corporation for Public Broadcasting college-level telecourse, *Against All Odds: Inside Statistics*, and for the series of video modules, *Statistics: Decisions through Data*, intended to aid the teaching of statistics in schools. He is the author of influential articles on statistics education and of several leading texts. Professor Moore has served as president of the International Association for Statistical Education and has received the Mathematical Association of America's national award for distinguished college or university teaching of mathematics.

George P. McCabe is Associate Dean for Academic Affairs in the College of Science and Professor of Statistics at Purdue University. In 1966, he received a BS degree in mathematics from Providence College and in 1970 a PhD in mathematical statistics from Columbia University. His entire professional career has been spent at Purdue,

with sabbaticals at Princeton University, the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Melbourne (Australia), the University of Berne (Switzerland), the National Institute of Standards and Technology (NIST) in Boulder, Colorado, and the National University of Ireland in Galway. Professor McCabe is an elected fellow of the American Association for the Advancement of Science and of the American Statistical Association; he was 1998 chair of its section on Statistical Consulting. In 2008–2010, he served on the Institute of Medicine Committee on Nutrition Standards for the National School Lunch and Breakfast Programs. He has served on the editorial boards of several statistics journals. He has consulted with many major corporations and has testified as an expert witness on the use of statistics in several cases.

Professor McCabe's research interests have focused on applications of statistics. Much of his recent work has focused on problems in nutrition, including nutrient requirements, calcium metabolism, and bone health. He is the author or coauthor of more than 190 publications in many different journals.

Bruce A. Craig is Professor of Statistics and Director of the Statistical Consulting Service at Purdue University. He received his BS in mathematics and economics from Washington University in St. Louis and his PhD in statistics from the University of Wisconsin–Madison. He is an elected fellow of the American Association for the Advancement of Science and of the American Statistical Association and was chair of its section on Statistical Consulting in 2009. He has also been an active member of the Eastern North American Region of the International Biometrics Society and was elected by the voting membership to the Regional Committee between 2003 and 2006.

Professor Craig has served on the editorial board of several statistical journals and has been a member of several data and safety monitoring boards, including Purdue's institutional review board.

Professor Craig's research interests focus on the development of novel statistical methodology to address research questions in the life sciences. Areas of current interest are diagnostic testing, inter-rater agreement, and abundance estimation. He is an author or

coauthor of more than 100 papers in more than 50 different journals.
In 2005, he was named Purdue University Faculty Scholar.

Data Table Index

TABLE 1.1	IQ test scores for 60 randomly chosen fifth-grade students
TABLE 1.2	Service times (seconds) for calls to a customer service center
TABLE 1.3	Educational data for 78 seventh-grade students
TABLE 2.1	Four data sets for exploring correlation and regression
TABLE 2.2	Two measures of glucose level in diabetics
TABLE 2.3	Dwelling permits, sales, and production for 21 countries
TABLE 2.4	World record times for the 10,000-meter run
TABLE 5.1	Length (in minutes) of 60 visits to a statistics help room
TABLE 7.1	Monthly rates of return on a portfolio (%)
TABLE 7.2	Parts measurements using optical software
TABLE 7.3	DRP scores for third-graders
TABLE 7.4	Seated systolic blood pressure (mm Hg)
TABLE	Length (in seconds) of audio files sampled from an iPod

7.5

TABLE Annual number of tornadoes in the United States between
10.1 1953 and 2014

TABLE In-state tuition and fees (in dollars) for 33 public
10.2 universities

TABLE Sales price and assessed value (in thousands of \$) of 35
10.3 homes in a midwestern city

TABLE Watershed area (km²), percent forest, and index of biotic
10.4 integrity

TABLE Iron content (mg/100 g) of food cooked in different pots
13.1

TABLE Tool diameter data
13.2

Beyond the Basics Index

Chapter 1	Density estimation
Chapter 2	Data mining
Chapter 3	Capture-recapture sampling
Chapter 4	More laws of large numbers
Chapter 5	Weibull distributions
Chapter 7	The bootstrap
Chapter 8	The plus four confidence interval for a single proportion
Chapter 8	The plus four confidence interval for a difference in proportions
Chapter 8	Relative risk
Chapter 9	Meta-analysis
Chapter 10	Nonlinear regression
Chapter 11	Multiple logistic regression
Chapter	Testing the equality of spread

CHAPTER 1

Looking at Data—Distributions



iStock/Creatista/Getty Images Plus

-
- 1.1 Data**
 - 1.2 Displaying Distributions with Graphs**
 - 1.3 Describing Distributions with Numbers**
 - 1.4 Density Curves and Normal Distributions**

Introduction

Statistics is the science of learning from data. Data are numerical or qualitative descriptions of the objects that we want to study. In this chapter, we will master the art of examining data.

We begin in [Section 1.1](#) with some basic ideas about data. We will learn about the different types of data that are collected and how data sets are organized.

[Section 1.2](#) starts our process of learning from data by looking at graphs. These visual displays give us a picture of the overall patterns in a set of data. We have excellent software tools that help us make these graphs. However, it takes a little experience and a lot of judgment to study the graphs carefully and to explain what they tell us about our data.

[Section 1.3](#) continues our process of learning from data by computing numerical summaries. These sets of numbers describe key characteristics of the patterns that we saw in our graphical summaries.

The final section in this chapter helps us make the transition from data summaries to statistical models that are used to draw conclusions and to make predictions. Specifically, we learn about using density curves to describe a set of data and are introduced to the Normal distributions. These distributions can be used to describe many sets of data that we will encounter. They also play a fundamental role in many of the methods of statistical analysis.

1.1 Data

When you complete this section, you will be able to:

- Give examples of cases in a data set.
- Identify the variables in a data set.
- Demonstrate how a label can be used as a variable in a data set.
- Identify the values of a variable.
- Classify variables as categorical or quantitative.
- Describe the key characteristics of a set of data.
- Explain how a rate is the result of adjusting one variable to create another.

A statistical analysis starts with a set of data. We construct a set of data by first deciding what *cases*, or units, we want to study. For each case, we record information about characteristics that we call *variables*.

CASES, LABELS, VARIABLES, AND VALUES

Cases are the objects described by a set of data. Cases may be customers, companies, subjects in a study, units in an experiment, or other objects.

A **label** is a special variable used in some data sets to distinguish the different cases.

A **variable** is a characteristic of a case.

Different cases can have different **values** of the variables.

EXAMPLE 1.1



COUPONS

Restaurant discount coupons. A website offers coupons that can be used to get discounts for various items at local restaurants. Coupons for food are very popular. [Figure 1.1](#) gives information for seven restaurant coupons that were available for a recent weekend. These are the cases. Data for each coupon are listed on a different line, and the first column has the coupons numbered from 1 to 7. The remaining columns give the type of restaurant, the name of the restaurant, the item being discounted, the regular price, and the discount price.



	A	B	C	D	E	F
1	ID	Type	Name	Item	RegPrice	DiscPrice
2	1	Italian	Domo's	Pizza	20	10
3	2	Italian	Mama Rita's	Pizza	20	12
4	3	BBQ	Smokey McSween's	Barbecue	30	17
5	4	BBQ	Smokey Grill	Ribs	20	11
6	5	Mexican	Dos Amigos	Tacos	16	8
7	6	Mexican	Holy Guacamole	Steak fajitas	13	8
8	7	Seafood	Sea Grille	Shrimp platter	20	11

FIGURE 1.1 Spreadsheet of food discount coupons, [Example 1.1](#).

Some variables, like the type of restaurant, the name of the restaurant, and the item simply place coupons into categories. The regular price and discount price columns have numerical values for

which we can do arithmetic. It makes sense to give an average of the regular prices, but it does not make sense to give an “average” type of restaurant. We can, however, do arithmetic to compare the regular prices classified by type of restaurant.

CATEGORICAL AND QUANTITATIVE VARIABLES

A **categorical variable** places a case into one of several groups or categories.

A **quantitative variable** takes numerical values for which arithmetic operations such as adding and averaging make sense.

EXAMPLE 1.2



COUPONS

Categorical and quantitative variables for coupons. The restaurant discount coupon file has six variables: coupon number, type of restaurant, name of restaurant, item, regular price, and discount price. The two price variables are quantitative variables. Coupon number, type of restaurant, name of restaurant, and item are categorical variables.

An appropriate label for your cases should be chosen carefully. In our food coupon example, a natural choice of a label would be the name of the restaurant. However, if there are two or more coupons available for a particular restaurant, or if a restaurant is a chain with

different discounts offered at different locations, then the name of the restaurant would not uniquely label each of the coupons. In the restaurant discount coupon file, the first variable, ID, is a unique label for each coupon.

spreadsheet

The display in [Figure 1.1](#) is from an Excel **spreadsheet**. Spreadsheets are very useful for doing the kind of simple computations that you will do in [Exercise 1.2](#). You can type in a formula and have the same computation performed for each row.

Note that the names we have chosen for the variables in our spreadsheet do not have spaces. For example, instead of “Restaurant Name” for the name of the restaurant, we simply use Name. *In some statistical software packages, however, spaces are not allowed in variable names.* For this reason, when creating spreadsheets for eventual use with statistical software, it is best to avoid spaces in variable names. Another convention is to use an underscore (_) where you would normally use a space. For our data set, we could have used Regular_Price and Discount_Price for the two price variables.



USE YOUR KNOWLEDGE

- 1.1 Read the spreadsheet.** Refer to [Figure 1.1](#). Give the regular price and the discount price for the Smokey Grill ribs coupon.
- 1.2 How much is the discount worth?** Refer to [Example 1.1](#). Consider adding another column to the spreadsheet that gives the coupon savings. Explain how you would compute the entries in this column. Does the new column contain values for a categorical variable or for a quantitative variable? Explain your answer.

unit of measurement

Another important part of the description of any quantitative variable is its unit of measurement. For both RegPrice and DiscPrice, the **unit of measurement** is clearly dollars. In other settings, it may not be as obvious. For example, if we were measuring heights of children, we might choose to use either inches or centimeters. The units of measurement are an important part of the description of a quantitative variable.